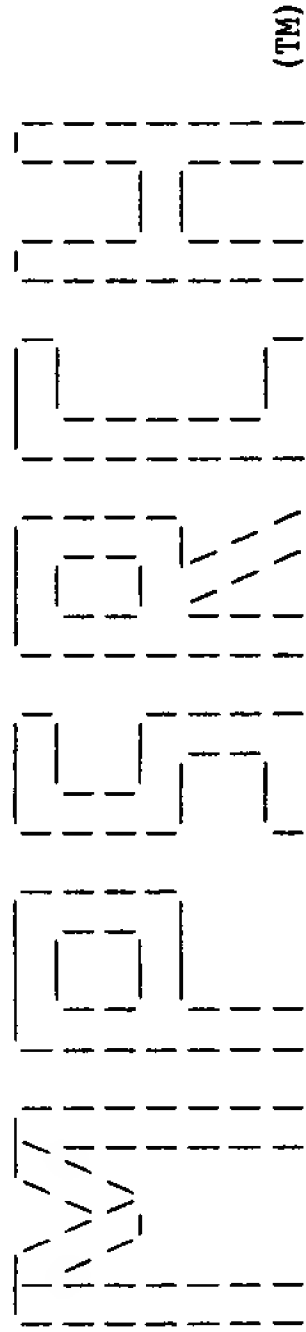


Jul9 17:02

US-08-717-294-42.rge

1



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Jul 9 15:13:00 1998; MasPar time 4882.54 Seconds
Tabular output not generated. 1521.186 Million cell updates/sec

Title: >US-08-717-294-42
Description: (1-4451) from US08717294.seq
Perfect Score: 4451
N.A. Sequence: 1 AACCTTAAACCATGCCCATG.....ACCTGTACTGAAGCGGCGGC 4451
Comp: TTCGAATTGGTACGGGTAC.....TGGACATGACTTCGCGCGCGC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0
Searched: 457396 seqs, 834335066 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl54
1:em_ba 2:em_htg 3:em_hum1 4:em_hum2 5:em_in 6:em_om
7:em_or 8:em_ov 9:em_pat 10:em_pl 11:em_ro 12:em_vi
Database: genbank106
13:gb_ba 14:gb_htg 15:gb_in 16:gb_om 17:gb_ov 18:gb_pat
19:gb_ph 20:gb_pl 21:gb_pr1 22:gb_pr2 23:gb_ro 24:gb_st
25:gb_sy 26:gb_un 27:gb_vi

Statistics: Mean 12.170; Variance 6.902; scale 1.763

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1612	36.2	4278	18	Sequence 4 from Patent	0.00e+00
2	1610	36.2	4281	18	Sequence 3 from Patent	0.00e+00
3	1087	24.4	6999	18	Sequence 1 from Patent	0.00e+00
4	1085	24.4	7272	18	Sequence 3 from Patent	0.00e+00
5	1085	24.4	8241	18	Synthetic Factor VIII	0.00e+00
6	1085	24.4	8241	18	Artificial mRNA for fa	0.00e+00

Jul9 17:02

US-08-717-294-42.rge

2

7	1085	24.4	8831	18	E00527	Complete cDNA encoding	0.00e+00
8	1085	24.4	9029	21	HUMFVIIIC	Human coagulation fact	0.00e+00
9	1083	24.3	7056	18	I27063	Sequence 1 from Patent	0.00e+00
10	1081	24.3	7440	18	I08345	Sequence 2 from Patent	0.00e+00
11	1083	24.3	8967	21	HSFVIIIR	Human mRNA for factor	0.00e+00
12	1083	24.3	8967	18	I71409	Sequence 1 from patent	0.00e+00
13	1083	24.3	9009	18	I31901	Sequence 3 from patent	0.00e+00
14	1083	24.3	9009	21	HUMFVIII	Human coagulation fact	0.00e+00
15	1083	24.3	9009	18	I63424	Sequence 1 from patent	0.00e+00
16	1073	24.1	7440	18	I08457	Sequence 1 from Patent	0.00e+00
17	1060	23.8	4548	18	I08642	Sequence 2 from Patent	0.00e+00
18	1061	23.8	4551	18	I08641	Sequence 1 from Patent	0.00e+00
19	1020	22.9	6539	16	SSU49517	Sus scrofa factor VIII	0.00e+00
20	1001	22.5	7032	16	AF016234	Canis familiaris facto	0.00e+00
21	991	22.3	7145	16	AF049489	Canis familiaris facto	0.00e+00
22	826	18.6	7493	18	I63427	Sequence 7 from patent	0.00e+00
23	826	18.6	7493	23	MUSCCVIII	Mus domesticus coagula	0.00e+00
24	758	17.0	1728	18	I02054	Sequence 8 from Patent	0.00e+00
25	664	14.9	1623	18	I27064	Sequence 3 from Patent	0.00e+00
26	606	13.6	1993	18	I02047	Sequence 1 from Patent	0.00e+00
27	496	11.1	3852	18	I04400	Sequence 35 from Paten	0.00e+00
28	495	11.1	3852	18	E00422	Human genomic DNA codi	0.00e+00
29	487	10.9	1130	18	I31900	Sequence 1 from patent	0.00e+00
30	487	10.9	1130	18	I63426	Sequence 5 from patent	0.00e+00
31	339	7.6	2493	21	HUMF8C	Homo sapiens coagulati	1.62e-216
32	310	7.0	955	18	E00526	Part of pESVDA vector	4.20e-195
33	242	5.4	1728	18	I02049	Sequence 3 from Patent	3.13e-145
34	142	3.2	5031	21	HUMFAC15	H.sapiens factor VIII	2.22e-73
35	129	2.9	1128	21	HUMFAC08	H.sapiens factor VIII	2.83e-64
36	123	2.8	391	21	HUMFAC12	H.sapiens factor VIII	4.21e-60
37	123	2.8	1728	18	I02052	Sequence 6 from Patent	4.21e-60
38	111	2.5	293	21	HUMFAC20	H.sapiens factor VIII	8.05e-52
39	112	2.5	669	18	I04388	Sequence 23 from Paten	1.66e-52
40	113	2.5	4074	21	HUMFAC13	H.sapiens factor VIII	3.41e-53
41	112	2.5	15155	18	I04391	Sequence 26 from Paten	1.66e-52
42	101	2.3	346	21	HUMFAC07	H.sapiens factor VIII	5.36e-45
43	102	2.3	2709	21	HUMFAC21	H.sapiens factor VIII	1.12e-45
44	96	2.2	218	21	HUMFAC17	H.sapiens factor VIII	1.29e-41
45	96	2.2	367	21	HUMFAC10	H.sapiens factor VIII	1.29e-41

ALIGNMENTS

RESULT	1	I08644	4278	bp	PAT	14-NOV-1994
LOCUS	DEFINITION	Sequence 4 from Patent	WO 8800831.			
ACCESSION	NID	I08644	g588650			
KEYWORDS	SOURCE	Unknown.				
ORGANISM	REFERENCE	1 (bases 1 to 4278)				
	AUTHORS	Pasek,M.P.				
TITLE		DNA SEQUENCES CODING FOR MODIFIED FACTOR VIII:C AND MODIFIED FACTOR VIII:C-LIKE POLYPEPTIDES AND PROCESSES FOR PRODUCING THESE POLYPEPTIDES IN HIGH YIELDS				
JOURNAL		Patent: WO 8800831-A 4 11-FEB-1988;				
FEATURES		Location/Qualifiers				
source		1. .4278				
BASE COUNT		1244 a 942 c 946 g 1146 t				
ORIGIN						

Qy	1464	CTGCTGATCATCTTCAAGAACGAGCCAGCAGCGCCCTACAACATCTACCCCCACGGCATC	1523
Db	1498	ACTGATGTCGCTCTTTCTATTCAAGGAGATTACCAAAAGGTGTAACAATTTGAAGCAT	1557
Qy	1524	ACCGACGTGGCCCCCTGTACAGCGCGCGCTGCCCAAGGGCGTGAAGACCTGAAGGAC	1583
Db	1558	TTTCCAAATTCTGCCAGGAGAAATATTCAAATATAAAATGGACACTGACTGTAGAAGATGGG	1617
Qy	1584	TTCCCCATCTCGCCGGCGAGATCTTCAAGTACAAGTGACCGCTGACCGTGGAGGACGGC	1643
Db	1618	CCAACTAAATCAGATCCTCGGTGCTGACCCGCTATTACTCTAGTTTTCGTTAATATGGAG	1677
Qy	1644	CCACCAAGAGGACCCCGCTGCTGACCCGCTACTACAGCAGCTTTCGTGAACATGGAG	1703
Db	1678	AGAGATCTAGCTTCAGGACTCATTTGGCCCTCTCCTCATCTGCTACAAAGAAATCTGTAGAT	1737
Qy	1704	CGCGACCTGGCCTCCGGACTGATCGCCCCCTGCTGATCTGCTACAGGAGCGCTGGAC	1763
Db	1738	CAAAGAGAAACACAGATAATGTCAACAGAGGAATGTCACTCTGTTTCTGTATTGTAT	1797
Qy	1764	CAGCGCGCAACAGATCATGAGCGACAAGCGCAACCTGATCTCTGTTACGCGTGTTCGAC	1823
Db	1798	GAGAACCGAAGCTGTACCTCACAGAGAAATATACAAAGCTTTCTCCCGAATCCAGCTGGA	1857
Qy	1824	GAGAACCGCAGCTGGTATCTGACCGAGAACATCCAGCGCTTCTCGCCCAACCCCGCTGGC	1883
Db	1858	GTGCAGCTTGAGCATCCAGAGTTCCAAGCCTCCAAGCATCATGCACAGGATCAATGGCTAT	1917
Qy	1884	GTGCAGCTGGAATCCCGAGTTCAGGCGCAGCAACATCATGCACAGCATCAACGGCTAC	1943
Db	1918	GTTTTTGATAGTTTGCAGTTGTCAGTTTGTTCGATCAGGTGGCATACTGGTACATTCTA	1977
Qy	1944	GTGTTCCAGAGCTGCAGCTGAGCGTGTGCTGCTCAGGTGGCTACTGGTACATCCTG	2003
Db	1978	AGCATTTGGAGCACAGACTGACTTCCTTTCTGTCTTCTCTGGAATATACCTTCAAACAC	2037
Qy	2004	AGCATCGCGCCAGACCGCACTTCTGAGCCTGTTCTTCTCCGGTATACCTTCAAGCAC	2063
Db	2038	AAATGGTCTATGAAGACACACTGACCCTATTCCCATTTCTCAGGAGAAACTGTCTTCATG	2097
Qy	2064	AAGATGCTGTACGAGGACACCCCTGACCCCTGTTCCTTCTCCTCCGGAGACTGTCTTCATG	2123
Db	2098	TGCATGGAACCCAGGCTATGGATTCTGGGTGCCACAACCTCAGACTTTCGGAACAGA	2157
Qy	2124	TCTATGGAGAACCCCGCCTGTGGATTCTGGGCTGCCACAACAGCGCACTTCGCGAACCGC	2183
Db	2158	GGCATGACCGCCCTTACTGAAGGTTTCTAGTTGTGACAAGAACACTGGTGATTATTACGAG	2217
Qy	2184	GGCATGACTGCCCTGCTGAAAGTCTCCAGCTGCCGACAAGAACCGGCGACTACTACGAG	2243
Db	2218	GACAGTTATGAAGATATTTACGACATACTTGCTGAGTAAACAAATGCCATTGAACCAAGA	2277
Qy	2244	GACAGCTACGAGGACATCTCCGCCCTACCTGCTGTCCAAGAACACAGCCCATCGAGCCCCG	2303
Db	2278	AGCTTCTCCAGNAATTCAGACACCCCTAGCACTAGGCAAAAGCAATTTAATGCCACC	2334
Qy	2304	TCCTTCTCCCAAAACTCCGCCACCCCGAGCAGCGGTGAGAAGCAGTTTCAACGCCACC	2360

RESULT5

LOCUSA053288241bpRNA

DEFINITIONSynthetic Factor VIII cDNA insert of pCLB89.

ACCESSIONA05328

NIDg345061

PAT03-MAY-1993

KEYWORDS	unidentified.
SOURCE	ORGANISM
REFERENCE	1 (bases 1 to 8241)
AUTHORS	
JOURNAL	Patent: WO 8809813-A 7 15-DEC-1988;
FEATURES	Location/Qualifiers
source	1..8241
CDS	/organism="Artificial sequences"
	7..7062
	/codon_start=1
	/product="Factor VIII"
	/db_xref="PID:g345062"
	/translation="MQIELSTCFFLCLRFCSATRRYYLGAVELSWDYMQSDLGELP
	VDARFPVRPKSFPFNTSVVYKTLFVEFTHLFINIAKPRPPWMGLLGPTIOAEVYDT
	VVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREXEDDKVFPGGSHTVVWQVL
	KENGPWASDPLCLITYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLF
	AVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHV
	IGMCTTPEVHSIFLEGHTFLVRNHRQASLEISPIFQLTAQTLMLDLQGLFLFCHISSH
	QHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDLTDSEMDVVRFDNNSPSFIQRS
	VAKHKPTWVHYIAAEEDWDYAPLVLPADDDRSYKSQYANNQPQRIGRKYKRVFMAY
	TDETFKTREAIQIHESGIPLLYGEVGDPLLIIFKNQASRPYNIYPHGIIDVRPLYSR
	RLPKGVKHLKDFIILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNMERDLASGL
	IGPLLI CYKESVDQRGNQIMSDKRNVI LFSVFENRSWYLTENIQRF LPNPAGVQLED
	PEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKKHWV
	YEDTLTFPFSGETVFMSENPGLMILGCHNSDFNRNRMWTALLKVSSCDKNTGDYYED
	SYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTPENDIEKTDPFWEAHTPM
	PKIQNVSSSDLLMLLRQSPTPHGLSLSDLQEAKEYETFSDDPSPGAIDSNNSLSEMTHF
	RPQLHSGDMVFTPESGIQLRLNEKLGTTAAATELKKLDLFKVSSTSNNLSTIPSDNLA
	AGTDNTSSLGPPSMPVHYDSQDITTLF GKSSP LTESGGPLS LSEENDSK LLESGLM
	NSQESSHGKNVSVSTESGRLEFKGRAHPALLTKDNALFKVSISLLKTNKTSNNSATNR
	KTHIDGPSLLIENSPSVMQNI L ESDTEFKKVTPLIHDRMLMDKNAFALRLNHSNKTT
	SSKNMEVQQKKEGP IPPDAQNPDMSF FKMFLPESARWIQRTHGKNSLNSGQPSPK
	QLVSLGPEKSVEGQNF LSEKNVYVVGKEFTKDVGLKEMVFPSSRNLF LTNLDNLHEN
	NTHNQEKKIQEEIEKKEETLIQENVVLPQIHTVTGTFNFMKNLFL LSTRQNVCEGSDGA
	YAPVLQDFRSINDSTNRTKKHTAHFSKKGEEENLEGNGNTKQI VEKVACTTRISPNT
	SQONFVTQRSKRALKQFRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNE
	KEKGAI TQSP LSDCLTRSHS IPOANRSP LPAKVSSFP SIRPI YLTRVLFQDNSSHLP
	AASYRKDQSGVQESSHFLQCAKKNLSLA ILTLEMTGDQREVGS LGTSATNSVTYKKV
	ENTVLPKPDLPKTSCKVELLPKVHIYQKDLFP TETSNGPSGHLDI VEGSLIQCTEGAI
	KWNEANRP GKVPFLRVATESSAKTPSKLJLDP LADNHYGTQIPKEEMKSQEKSPKTA
	FKKKDTILSINACESNHAIAA INEQNKPEIEV TWAKQGRTERLC SQNP PVLKRHQRE
	ITRTTLQSDQEEIDYDDTISVEMKKEDFIYDEDENQSPRSFQKKTRHYF IAAVERLW
	DYGMSSSPHVLNRNAQSGVPQFKVFOEFTDGSFTQPL YRGEI NEHGLLGPYIRA
	EVEDNIMVTFRNQASRPYSFYSS LISYEEDQRGAEPKRFVKPNE TKTYFKVQVHEM
	APT KDEFDCKAWAYFSDVDLEKDVHSG LIGPLLVCHTNTLNPAHGRQVTVQEFALFFT
	IFDETKSWYFTENMERNCRAPCN IQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQR
	IRWYLLSMGSNENIHSIHFSGHVF TVRKKEEYKMA LNYLPGVFETVEMLP SKAGIWR
	VECLIGEHLHAGMSTLFLVYSNKQTP LGMASGHIRDFOITASGQYQGWAPKLARLHY
	SGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKQSSLYISQFI MYSLDGKKWQ
	TYRGNSTGTLMVFFGNVDSSGIKHNIFNPP I IARYIRLHP THYSIRSTLRMEWMGCDL
	NSCSMP LGMESKAI SDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEW
	LQVDFQKTMKVTVGTTQGVKSLLTSMYKFE LSSSQDGHQWTLFFQNGKVKVFGNQ
	DSFTPVVNSLDPP LLLTRYLRIRHPQSVWHQIALRMEVLGCEAQDLY"
BASE COUNT	2625 a 1759 c 1703 g 2154 t
ORIGIN	

Query Match24.4%; Score 1085; DB 18; Length 8241;

Best Local Similarity73.8%; Pred. No. 0.00e+00;

Matches 1681; Conservative0; Mismatches 596; Indels0; Gaps0;

Db64GCCACCAGAAAGTACTACCTGGTGCAGTGGAAGTGTCTATGGGACTATATGCAAGTGAT123

||||| | | ||||| ||||| | |||| | |||| | ||||| ||||| ||||| ||||| ||

Qy 84 GCCACCGCCGCTACTACCTGGCGCCGCTGCGAGCTCTCTCTGGGACTACATGCAGCGAC 143

Db 124 CTCGGTGAGCTGCCTGTGGACGGAAAGATTTCCTCCTAGAGTGGCAAATCTTTTCCATTTC 183
|| || |||| || ||||| || || || || || ||||| || || || || ||

Qy 144 CTGGGCGAGCTCCCGTGGAGCGCCGCTTCCCCCCGCGGTGCCCAAGAGCTTCCCTTC 203

Db 184 AACACCTCAGTCGTGTACAAAAAGACTCTGTTGTAGAAATTACGGGATCACCTTTTCAAC 243
||||| || ||||| || || ||||| || || ||||| || || ||||| |||||

Qy 204 AACACCGCGTGTGTACAAGAAAACCCCTGTCTGTGAGTTACCGACCACTTTCACAC 263

Db 244 ATCGCTAAGCCAGGCGCCCTTGGATGGGTCTGCTAGGTCCTACCATCCAGGCTGAGGTT 303
|| || |||| || || ||||| || || ||||| || || ||||| || || ||||| |||||

Qy 264 ATTGCCAAGCCGGCCCCCTGGATGGGCTGTCTGGGCCCCACCATCCAGGCGAGTG 323

Db 304 TATGATACAGTCGTCAATTACACTTAAAGAACATGGCTTCCCATCCTGTCACTCTTCATGCT 363
|| || || ||||| || || || ||||| || || || ||||| || || || || || || ||

Qy 324 TACGACACCGTGTGTACCCCTGAAGAACATGGCCAGCCACCCGTCAGCCTGCAGGCC 383

Db 364 GTTGGTGTATCCTACTCTGAAAGCTTCTGAGGGAGCTGAATATGATGATCAGACAGTCAA 423
|| || || ||||| || || ||||| || || || || || || || || || || || || ||

Qy 384 GTGGGCTGAGCTACTGGAAGGCCAGCGAGGGCGCGGAGTAGCAGACAGACTCCCGAG 443

Db 424 AGGGAAGAAAGATGATAAAGCTTCTCCCTGGTGGAGGCCATACATATGCTGGCAGGTC 483
|| |||| || || || || || ||||| || || || || || ||||| || || || ||||| ||

Qy 444 CGCGAGAAGGAGCAGACAAGGTCTCCCGGGGGAGCCACACCTACGTTGTGCAGSTG 503

Db 484 CTGAAGAGAAATGGTCCAAATGGCCTCTGACCCACTGTGCCCTTACCTACTCATATCTTTCT 543
|| || |||| || || ||||| || || || ||||| || || || ||||| || || ||

Qy 504 CTTAAGGAGAACGGCCCCATGGCCAGCGACCCCCCTGTGCCCTGACCTACAGCTACCTGAGC 563

Db 544 CATGTGGACCTGGTAAAGACTTGAATTCAGGCCCCTCATTGGAGCCCCTACTAGTATGTAGA 603
|| ||||| ||||| || || |||| || || || || || || || || || || || || || ||

Qy 564 CACGTGGACCTGGTGAAGGATCTGAACAGCGGCTGATCGGCGCCCTGCTGCTGTGCGC 623

Db 604 GAAGGAGCTCTGCCCAAGGAAAGACACAGACACTTGCACAAATTTATACTACTTTTGCT 663
|| || || ||||| ||||| || || ||||| ||||| || || ||||| || || || || || || ||

Qy 624 GAGGGAGCCTGSCCAAGGAGAAACCCAGACCCCTGACAACTTCATCCTGCTTTCGCC 683

Db 664 GTATTTGATGAGGGAAAGTTGGCACTCAGAAACAAAGAACTCCTTGATGCAGGATAGG 723
|| || || || ||||| || ||||| || || ||||| || || ||||| || || || || || || ||

Qy 684 GTGTTCCAGCAGGGAGAGCTGGCACAGCGAGACTAAGAACAGCCTGATGCAGGACCGC 743

Db 724 GATGCTGCATCTGCTCGGSCCTGGCTAAATGCACACAGCTCAATGGTTATGTAAACAGG 783
|| || || || || ||||| || || ||||| || || ||||| || || || || || || || ||

Qy 744 GACGCGCCAGCGCCGCGCCTGGCCCCAAGATGCACACCGTTAACGGCTACGTCAACCGC 803

Db 784 TCTCTGCCAGGTCTGATGGATGCCACAGGAAATCAGTCTATTGGCATGTGATGGAAATG 843
||||| || ||||| || ||||| || || || || || || || || || || || || || || || || || ||

Qy 804 AGCCTGCCGGCTGATCGGCTGCCACCGCAAGGCGTGTACTGGCACGTCATCGGCAATG 863

Db 844 GGCACCACTCCTGAAGTGCACTCAATATTCTCGAAGGTCACACATTTCTTTGAGGAAC 903
|||||| ||||| ||||| || ||||| || || ||||| || || ||||| || || || || || || || ||

Qy 864 GGCACACCCCTGAGGTGCACAGCATCTTCCTGGAGGGCCACACCTTCTGTTGGTGGCAAC 923

Db 904 CATCGCCAGGCGCTTGGAAATCTCGGCCAATAACTTTTCTTACTGCTCAAAACACTCTTG 963
|| ||||| || || ||||| || || || || || || || || || || || || || || || || || ||

Qy 924 CACCGCAGGCGCAGCTGGAGATCAGCCCCATCACCCTTCCTCACTGCCAGACCCCTGCTG 983

Db 964 ATGGACCTTGGACAGTTTCTACTGTTTGTCAATATCTCTTCCCACCAACATGATGGCAATG 1023
|||||| || || || || || || || || || || || || || || || || || || || || || || ||

Qy 984 ATGGACCTAGGCCAGTTTCTGCTGTCTGCCACATCAGCAGGCCACCAAGAGGGCAATG 1043

Db 1024 GAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCAACTACGAAATGAAAAATAAT 1083
|| ||||| || || || ||||| || || ||||| || || ||||| || || ||||| || || || || || || ||

Qy 1044 GAGGCTTACGTCAAGGTGGACAGCTGCCCGGAGGAGCCCCAGCTGCCCATGAAGAACAAAC 1103

Db 1084 GAAGAAGCGGAAGACTATGATGATCTTACTGATTTCTGAATGGAATGCTGCTCAGGTTT 1143
|| || || || ||||| || || || || || || || || || ||||| || || || || ||

Qy 1104 GAGGAGCGGAGGACTAGGACGACGACCTTGACCGGACAGCGAGATGATGTCGTACGCTTC 1163

Db 1144 GATGATGACAACTCTCCTTCCTTTATCCAAATTCGCTCAGTTGCCAAGAAGCATCCTAAA 1203
|| || ||||| || || ||||| || || ||||| || || || ||||| |||||

Qy 1164 GACGACGACAACGCCCCAGCTTCATCCAGATCCGACGCTGGCCAAAGAAGCACCCCTAAG 1223

Db 1204 ACTTGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTAGTCCTC 1263
|| ||||| || ||||| || || || ||||| ||||| || || ||||| || || || || || || ||

Qy 1224 ACCTGGGTGCACTACATCGCGCGGAGGAGGAGGACTGGGACTACGCCCGGTAGTACTG 1283

Db 1264 GCCCCGATGACAGAACTTATAAAAGTCAATATTTGAACAAATGGCCCTCAGCGGATTGGT 1323
|||||| || || || || || || || || || || ||||| ||||| ||||| || || || || ||

Qy 1284 GCCCCGACGACCGCAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGCGCATCGGC 1343

Db 1324 AGGAAGTACAAAAAGTCCGATTATAGGATACAGATGAAAACCTTTAAGACTCGTGAA 1383
|| ||||| || || || || || ||||| || || || || || || || || || || || || ||

Qy 1344 CGCAAGTACAAGAAGGTGCGCTTCATGGCTACACCGACGAGACTTCAAGACCCGCGAG 1403

Db 1384 GCTATTCAAGCATGAATCAGGAATCTTGGACCTTTACTTTATGGGGAAGTTGGAGACACA 1443
|| || ||||| || || || || || || || || || || || || || || || || || || || || || || || || ||

Qy 1404 GCCATCCAGCAGGAGTCCGGCATCCTCGGCCCTGCTGACGGCGAGGTGGCGACACC 1463

Db 1444 CTGTTGATTTATATTTAAGAATCAAGCAACGACAGACCATATAACATCTACCCCTACCGGAATC 1503
|| || ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || || || || || || ||

Qy 1464 CTGCTGATCATCTTCAAGAACCCAGGCCAGCAGSGCCTACAACATCTACCCCAACGGCATC 1523

Db 1504 ACTGATGTCCGTCCCTTTGTATTTCAAGGAGATTACCAAAAGGTGTAAAACATTTGAAGGAT 1563
|| || || || || || || || || || || || || || || || || || || || || || || || || || || || ||

Qy 1524 ACCGAGTGGCGCCCTGTACAGCGCCGCTGCCAAGGCGTGAAGCACCTGAAGGAC 1583

Db 1564 TTTCCAAATTTGCCAGGAGAAATATTCAAATATAAATGGACAGTGAAGTGAAGATGGG 1623
|| || || ||||| || || || ||||| || || ||||| || || ||||| || || || || || || || ||

Qy 1584 TTCCCATCCTGCCCCGGAGATCTTCAAGTACAAGTGGACCGTGACCGTGGAGGACGGC 1643

Db 1624 CCAACTAAATCAGATCCTCGGTGCCTGACCCGCTATTACTCTAGTTTCGTTAAATATGGAG 1683
|| || || || || || || || || ||||| ||||| || || ||||| || || || || || || || || || || || || ||

Qy 1644 CCCACCAAGAGCGACCCCGCTGCCCTGACCCGCTACTACAGCAGCTTCGTGAACATGGAG 1703

Db 1684 AGAGATCTAGCTTCAGGACTCATTTGGCCCTCTCCTCATCTGCTACAAAGAATCTGTAGAT 1743
|| || || || || ||||| || ||||| || || ||||| ||||| || || ||||| || || || || || || || ||

Qy 1704 CGGACCTGGCTCCGACTGATCGGCCCCCTGCTGATCTGCTACAAGGAGAGCGTGGAC 1763

Db 1744 CAAAGGGAACCCAGATAATGTCAGACAAAGGAATGTCATCCTGTTTCTGTATTTGAT 1803
|| || || ||||| || || ||||| || || ||||| || || ||||| || || || || || || || || || || || || ||

Qy 1764 CAGCGGGCAACCCAGATCATGCGGACAAGCGCAACGCTGATCCTGTTCAAGGAGAGCGTTCGAC 1823

Db 1804 GAGAACCGAAGCTGGTACCTCAGACAGAAATATACAACGCTTCTCCCCAATCCAGCTGGA 1863
|||||| ||||| ||||| || || ||||| || || ||||| || || ||||| || || || || || || || || || || || || ||

Qy 1824 GAGAACCGCAGCTGGTATCTGACCGAGAACATCCAGCGCTTCTGCCCAACCCCGCTGGC 1883

Db 1864 GTGACCTTGAGGATCCAGAGTTCCCAAGCCTCCAACATCATGCACAGCATCAATGGCTAT 1923
|||||| || ||||| || ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1884 GTGACCTGGAAGATCCCGAGTTCAGGCGGCAACATCATGCACAGCATCAACGGCTAC 1943

Db 1924 GTTTTGTATAGTTTGCAGTTGTCAGTTTGTGATGAGGTGGCATACTGGTACATTTCTA 1983
|| || || || ||||| || || || ||||| || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1944 GTGTTGCACAGCCTGCAGCTGAGCGTGTGCCTGCATGAGTGGCCTACTGTTACATCTCTG 2003

Db 1984 AGCAATTGAGCACAGACTGACTTCCCTTTCTGTCTTCTCTGGATATACCTTCAAACAC 2043
||||| || || || ||||| ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || || || || || ||

Qy	2004	AGCATCGCGGCCACGCGACTTCCTGAGCGGTTCCTCTCGGGTATACCTTCAAGCAC	2063
Db	2044	AAAAATGGTCTATGAAGACACACTACCCCTATTCCCAATTCTCAGAGAAACTGTCTTCATG	2103
Qy	2064	AAGATGGGTACGAGGACACCCCTGACCCCTGTCCCCCTTCTCCGCGGAGACTGTGTTCATG	2123
Db	2104	TCGATGGAAAAACCCAGGCTTATGGATTCTTGGGTTGCCACAACACTCAGACATTTCGGAACAGA	2163
Qy	2124	TCATGGAACAACCCGGCCTGTGGATTCTGGCTGCCACAACACGCGACTTCGCGCAACCGC	2183
Db	2164	GGCATGACCGCCCTTACTGAAGGTTTCTAGTTGTGACAAACAACACTGGTGATTATTACGAG	2223
Qy	2184	GGCATGACTGCCCTGCTGAAAGTCTCCAGCTCGGACAACAACACCGGGGACTACTACGAG	2243
Db	2224	GACAGTTATGAAGATATTTTCAGCATACTTGCCTGAGTAAACAATGCCATTGAACCAAGA	2283
Qy	2244	GACAGCTACGAGGACATCTCGGCCTACCTGCTGCTCCAAGAACAACGCCATCGAGCCCCGC	2303
Db	2284	AGTTTCTCCCAAGATTCAACACACCCCTAGCACTAGGCAAAAGCAATTTAATGCCACC	2340
Qy	2304	TCCTTCTCCAAAACCTCCGCCACCCACGACCGCTCAGAAGCAGTTCAACGCCACC	2360
RESULT	6		
LOCUS	A07042	8241 bp	RNA
DEFINITION	Artificial mRNA for factor VIII.	PAT	24-AUG-1993
ACCESSION	A07042		
NID	g412967		
KEYWORDS	factor VIII.		
SOURCE	unidentified.		
ORGANISM	unidentified		
	unclassified.		
REFERENCE	1	(bases 1 to 8241)	
AUTHORS	Van Ooyen,A.J.J., Andreoli,P.M., Van Mourik,J.A. and Pannekoek,H.		
TITLE	Method for the preparation of proteins with factor VIII activity by microbial host cells; expression vectors, host cells, antibodies		
JOURNAL	Patent: EP 0253455-A 23 20-JAN-1988;		
	GIST-BROCADES N.V		

FEATURES	Location/Qualifiers
source	1..8241
	/organism="Artificial sequences"
CDS	7..7062
	/codon_start=1
	/product="factor VIII"
	/db_xref="PID:g412968"
	/translation="MQIELSTCFFLCLLRFCFSATRRYYLGVAVELSWDMQSDLGELP
	VDAREPPRPVKSPFNTSVYKKTLFVEETHLENIAKPRPPWMLGLGPTIQAEVYDT
	VVITLKNMASHPVSLHAGVSVYKASEGEYDDQTSQREKEDDKVFPGGSHTYVQVL
	KENGPMA SDP LCLTYSYLSHYDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLF
	AVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRSVYVHV
	IGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTLLMDLGQFLIFCHISSH
	QHDGMEA YVKVDSCEEPQLRMKNNEAEYDDDLTDSEMDVVRFDNDNSPSFIOIRS
	VAKKHPKTVVHYIAAEEDWDYAPLVLPADDRYSQSYLNNGPQRIGRKYKKVREWAY
	TDFTFKTREA IQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHIGTIDVRPLYSR
	RLPKGVKHLKDFPILPGEIFYKWTVTVEDGPTKSDPRLTRYSSFFVNERDLASGL
	IGPLLICYKESVDQRGNQIMSDKRNVLFSVDENRSHWLTENIQRELNPAGVQLED
	PEFQASINMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDLSVFFSGYTFKHKV
	YEDTLTLPFFSGETVFMSEMPNGLWILGCHNSDFRNRMGTALLKVVSSCDKNTGDYED
	SYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTIPENDIEKTDPFWAHRTPM
	PKIQNVSSDDLMLLRQSPTHGLSLDLQEAKEYETFSDPSPGAIDSNNSLSEMTFH
	RPQLHHS GDMVFTPESGLQRLINKEKLTAAATELKKLDFKVSSTNNLITIPSDNLA
	AGTDNTSSIGPPSMPVHYDSQDITTLFGKSSPLTESGFLSLEENND SKLLESGLM
	NSQESSWGKNVNSTESGRLFKGRAGPALLTKDNALFKVISILKTKNTSNNSATNR
	KTHIDGPSLLIENSPSVWQNTILESDTFEFKVTPLIHDRMLMKNATAIRLNHMSNKT

SSKNMEMVQQKKEGP IPPDAQNPDMSEFFKMLFLESARNIQRTHGKNSLNSGQPSPK
QILVSLGPEKSVEGONFLSEKNKVVGKGEFTKVGILKEMVFPSSRNLF LTNLDNLHEN
NTHNQEKKIQEEIEKKETLIQENVVLPQIHTVTGTFKNFMKNLFLLSTRQNVESGYDGA
YAPVLQDFERSLNDSTNRTKKHTAHFSKKGEENLEGLNQTKQIVEKYACTTRISPNT
SQQNFVTRSKRAALKQFRILPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNE
KEKGAITQSP LSDCLTRSHSIPOANRSLPIAKVSSFPSIRPIYLTRVLFDQNSSHLP
AASYRKDQSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGLS LGSATNSVTYKKV
ENTVLPKDPDLKPTSGKVELLPKVHIIYQKDLFPTETSNQSPCHLDLVEGSLLOQTEGAI
KWNEANRP GKVPFLRVATESSAKTPSKLDP LADNHYGTQIPKEEMKWSQEKSPKTA
FKKDOTILSLNACESNHAIAAINEGQNKPEIEVTWAKQGRTERLCSQNPPV LKRHQRE
ITRTTLQSDQEEIDYDDTISVEMKDEFDIYDEDENQSPRSFQKKTRHYFIAAVERLW
DYGMSSSPHVLNRNRAQSGVPQFKKVVFQEETDGSFTQPILYRGELNEHLGLLGPYIRA
EVEDNIMVTRNQAASRPYSFYSSLSI SYEEDQRGAEPKRFVKPNETKTYFWKVQHIM
APTQDEFDCKAWAYFSDVDLEKDVHSGLIPLLVCHTNTLNPAGHQVTVQEFALFFT
IFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDOR
IRWYLLSMGNSENIHSIHFSCHVFTVRKKEEYKMALYNLYPGVFETVEMLP SKAGIWR
VECLIGEHLHAGMSTLFLVYSNKCQTPLGMA SGHIRDFQITASGQYQWAPKILARLHY
SGSINAWSTKEPFSWIKVDLLAPMIIHGIKTGARQKFSLSIYSQFTIMYSLDGKKWQ
TYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLPHTHYSIRSTLRMEMMGCDL
NSCSMPLGMESKAI SDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKW
LQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFFQGNQ
DSFTPVVNSLDPP LLLTRYLRIRHPQSWVHQIALRMEVLGCEAQDLY*

BASE COUNT	2625 a	1759 c	1703 g	2154 t
ORIGIN				
	Query Match	24.4%;	Score 1085;	DB 18; Length 8241;
	Best Local Similarity	73.8%;	Pred. No. 0.00e+00;	
	Matches 1681;	Conservative 0;	Mismatches 596;	Indels 0; Gaps 0;
Db	64	GCCACCAGAAATACTACCTGGTGCAGTGGAACTGTCTATGGGACTATATGCAAAAGTCAT	123	
Qy	84	GCCACCCGCCGTACTACTGCGCGCGCTGGAGCTGTCTCTGGGACTACATGCAGAGCCAC	143	
Db	124	CTCGGTGAGCTGCCTGTGGACGCAAGATTTCCTCTAGAGTGCCTAAATCTTTTCCATTC	183	
Qy	144	CTGGGCGAGCTCCCCGTGACGCGCCGCTTCCCCCCCCCGCTGCCCAAGAGCTTCCCCCTTC	203	
Db	184	AACACCTCAGTCGTGTACAAAAGACTCTGTTTGTAGAAATCACGGATCACCTTTTCAAC	243	
Qy	204	AACACCGCGTGTGTACAGAAAACCCCTGTTCGTGGAGTTCACCGACCACCTGTTCAAC	263	
Db	244	ATCGCTAAGCCAGGCCACCTCGATGGGTCTGTAGTCTCTACCATCCAGGCTGAGCTT	303	
Qy	264	ATTGCCAAGCGCGCGCCCTCGATGGGCTGTCTGGGCCCCACCATCCAGGCCGAGGTG	323	
Db	304	TATGATACAGTGGTCACTACACTTAAGAACATGGCTTCCCATCTGTCTCAGTCTTTCATGCT	363	
Qy	324	TACGACACCGTGTGTATCACCCCTGAAGAACATGGCCAGCCACCCCGCTCAGCTGCACGCC	383	
Db	364	GTTGCTGTATCCTACTCGAAAAGCTTCTGAGGGAGCTGAATATGATGATCAGACCACTCAA	423	
Qy	384	GTGGCGCTGAGTACTGGAAGGCCAGGAGGGCGCGGAGTACGACGACCGACGCTCCAG	443	
Db	424	AGGAGAAAGAGATGATAAAGTCTTCCTGTTGGAAGCCATACATATGTGTGGCAGTTC	483	
Qy	444	CCCGAGAGGAGGACGACAAGGTGTTCCCGGGGGGAGGCACACCTACCTGTGGCAGGTG	503	
Db	484	CTGAAAGAGAAATGGTCCAAATGGCTCTGACCCACTGTGCCTTACCTACTCATATCTTTCT	543	
Qy	504	CTTAAGGAGAACGGCCCTATGGCCAGCGACCCCTCTGTGCTGACCTACAGTACCTGAGC	563	
Db	544	CATGTGCACCTGTTAAAGACTTGAATTCAGGCCTCATTTGGAGCCCTACTAGTATGTAGA	603	
Qy	564	CACGTGGACCTGTTGAAGGATCTGAACACGCGGGCTGATCGCGGCCCTGCTGCTGTCTCGC	623	

Db	604	GAAGGAGTCTGGCCAAAGGAAACACACACACCTTGACACAAATTTATACTACTTTTTGCT	663
Qy	624	GAGGCGAGCTGGCCAAAGGAGAAACCCAGACCCCTGCACAAGTTCATCCTGCTGTTGGCC	683
Db	664	GTATTTGATGAAGGGAAGAAAGTTGGCACTCAGAAACAAAGAACTCCTTGATGCAGGATAGG	723
Qy	684	GTCTTCGACGAGGGGAAGAGCTGCACACGCGAGACTAAGAACAGCCTGATGCAGGACCGC	743
Db	724	GATGCTGCATCTGCTCGGCGCCTGGCCTAAATGCACACAGTCAATGGTTATGTAACACAGG	783
Qy	744	GACGCGCCAGCGCCCGCGCCTGSCCCCAAGATGCACACCGTTAAGCGGTACGTGAACCGC	803
Db	784	TCTCTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATG	843
Qy	804	AGCCTGCCCGCCTGATCGGCTGCCACCCGCAAGAGCGTGTACTGGCACCTCATCGGCATG	863
Db	844	GGCACCACTCTGAAATGCACCTCAATATTTCTCGAAGGTACACATTTCTTTGTGAGGAAC	903
Qy	864	GGCACCACTCTGAAATGCACCTCAATATTTCTCGAAGGTACACATTTCTTTGTGAGGAAC	923
Db	904	CATCGCCAGCGCTCTTGGAAATCTCGGCAATAACTTTCTTACTGCTCAAAACACTCTTG	963
Qy	924	CACCGCCAGCGCTCTTGGAAATCTCGGCAATAACTTTCTTACTGCTCAAAACACTCTTG	983
Db	964	ATGACCTTTGGACAGTTTCTACTGTTTGTTCATATCTCTTCCACCAACATGATGGCATG	1023
Qy	984	ATGACCTTTGGACAGTTTCTACTGTTTGTTCATATCTCTTCCACCAACATGATGGCATG	1043
Db	1024	GAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCACTACGAATGAATAATAAT	1083
Qy	1044	GAGGCTTACGTGAAGGTGACAGTGTCCCGAGGAGCCCGAGTGGCATGGCATGAAGAACAAC	1103
Db	1084	GAAGAAGCGGAAGACTATGATGATCTTACTGATTTCTGAATGGATGTTGGTACGGTTT	1143
Qy	1104	GAGGAGCGCGAGACTACGACGACGACCTGCACCGACAGCGAGATGGATGCTGACGCTTC	1163
Db	1144	GATGATGACAACTCTCCTTCTTTATCCAAATTCGGTCAGTTGCCAAGAGCATCCTAAA	1203
Qy	1164	GACGACGACAAACAGCCCGAGCTTTCATCCAGATCCGAGCGTGGCCCAAGAGCACCTAAG	1223
Db	1204	ACTTGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTAGTCCCTC	1263
Qy	1224	ACCTGGGTGACACTACATCGCCGCGCAGGAGGAGGACTGGGACTACGCCCGGCTAGTACTG	1283
Db	1264	GCCCCCGATACAGAACTTATAAAGTCAATATTTGAACAATGGCCCTCAGCGGATTTGGT	1323
Qy	1284	GCCCCCGACGACCGGAGCTACAGAGCGGAGTACCTGAACAACAGGCCCCCAGCGCATCGGC	1343
Db	1324	AGGAAGTACAAAAAGTCCGATTTATGGCATACACAGATGAACCTTTAAGACTCCTGAA	1383
Qy	1344	CGAAAGTACAAGAAAGTGGCTTCATGGCTACACCGACGACACTTTTCAAGACCCCGGAG	1403
Db	1384	GCTATTACGATGAATCAGGAATCTTTGGACCTTTTACTTTATGGGGAAGTTGGACACACA	1443
Qy	1404	GCCATCCAGCAGAGTCCCGGATCTCTCGGCCCCCTGCTGTACGGCGAGGTGGCGACACC	1463
Db	1444	CTGTTGATTATATTTAAGAATCAAGCAAGCAGACCATATACATCTACCTTCACGGAATC	1503
Qy	1464	CTGCTGATCATCTTCAAGAACCAGGGCCAGCAGGCGCTTCAACATCTACCCCCACGGCATC	1523
Db	1504	ACTGATGTCGCTCTTTGATTCAAGGAGATTACCAAAAGGTCTAAACATTTGAGGAT	1563
Qy	1524	ACCGACGTGGCCCCCTGTACAGCCCGCGCTGCCCAAGGCGGTGAAGCACCTGAAGGAC	1583

Db	1564	TTTCCAAATTCTGCCAGGAGAAATATTCAAATATAAAATGGACAGTGA	1621	TTTCCAAATTCTGCCAGGAGAAATATTCAAATATAAAATGGACAGTGA
Qy	1584	TTCCCCATCTCTGCCCGCGAGATCTTCAAGTACRAAGTGGACCGTGACCGTGGAGGACGGC	1643	TTCCCCATCTCTGCCCGCGAGATCTTCAAGTACRAAGTGGACCGTGACCGTGGAGGACGGC
Db	1624	CCAACATAATCAGATCCTCGTGCCTGACCCGCTATTACTTAGTTTCGTTAATATGGAG	1683	CCAACATAATCAGATCCTCGTGCCTGACCCGCTATTACTTAGTTTCGTTAATATGGAG
Qy	1644	CCCACCAAGAGCGACCCCGCTGCCTGACCCGCTACTACAGCAGCTTCGTTGAACATGGAG	1703	CCCACCAAGAGCGACCCCGCTGCCTGACCCGCTACTACAGCAGCTTCGTTGAACATGGAG
Db	1684	AGAGATCTAGCTTCAGGACTCATTTGGCCCTCTCTCATCTGCTACAAAGAACTGTGTAGAT	1743	AGAGATCTAGCTTCAGGACTCATTTGGCCCTCTCTCATCTGCTACAAAGAACTGTGTAGAT
Qy	1704	CGCGACCTGGCCTCCGGACTGATCGGCCCTGCTGATCTGCTACAGGAGAGCGGTGGAC	1763	CGCGACCTGGCCTCCGGACTGATCGGCCCTGCTGATCTGCTACAGGAGAGCGGTGGAC
Db	1744	CAAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCTATCCTGTTTCTGTATTGTAT	1803	CAAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCTATCCTGTTTCTGTATTGTAT
Qy	1764	CAGCGCGGCAACCATCATCAGCGACAGCGCACCGTATCCTGTTTCAGCGGTGTTTCGAC	1823	CAGCGCGGCAACCATCATCAGCGACAGCGCACCGTATCCTGTTTCAGCGGTGTTTCGAC
Db	1804	GAGAACCGAAGCTGTGTACCTCACAGAGAAATATACAACGCTTTCTCCCAATCCAGCTGGA	1863	GAGAACCGAAGCTGTGTACCTCACAGAGAAATATACAACGCTTTCTCCCAATCCAGCTGGA
Qy	1824	GAGAACCGAGCTGTATCTCAGCCGAGAACATCCAGCGCTTCCTGCCCAACCCCGCTGGC	1883	GAGAACCGAGCTGTATCTCAGCCGAGAACATCCAGCGCTTCCTGCCCAACCCCGCTGGC
Db	1864	GTGAGCTTGAGATCCAGAGTTCGAAGCCTCCAACATCATGCACAGATCAATGGCTAT	1923	GTGAGCTTGAGATCCAGAGTTCGAAGCCTCCAACATCATGCACAGATCAATGGCTAT
Qy	1884	GTCAGCTTGAAGATCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGCTAC	1943	GTCAGCTTGAAGATCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGCTAC
Db	1924	GTTTTTGATGATTTGCAGTTGTCAGTTGTTTGGATGAGGAGGCACTGGTACATTTCTA	1983	GTTTTTGATGATTTGCAGTTGTCAGTTGTTTGGATGAGGAGGCACTGGTACATTTCTA
Qy	1944	GTTGTCAGACGCTGCAGCTGAGCGTGTGCCTGCATGAGGTGGCTACTGGTACATCCTG	2003	GTTGTCAGACGCTGCAGCTGAGCGTGTGCCTGCATGAGGTGGCTACTGGTACATCCTG
Db	1984	AGCATTTGGAGCACAGACTGACTTCTCTTTCTGTCTTCTCTGGATATACCTTCAAACAC	2043	AGCATTTGGAGCACAGACTGACTTCTCTTTCTGTCTTCTCTGGATATACCTTCAAACAC
Qy	2004	AGCATGGCGCCAGACCGACTTCTTGAGCGTGTCTTCTCCGGGTATACCTTCAAAGCAC	2063	AGCATGGCGCCAGACCGACTTCTTGAGCGTGTCTTCTCCGGGTATACCTTCAAAGCAC
Db	2044	AAATGGTCTATGAAGACACACTCACCTATTTCCCATTTCTCAGGAGAAACTGTCTTCATG	2103	AAATGGTCTATGAAGACACACTCACCTATTTCCCATTTCTCAGGAGAAACTGTCTTCATG
Qy	2064	AAGATGGTGTACGAGGACACCTGACCTGTTCCTTCTCCGGGTATACCTTCAAAGCAC	2123	AAGATGGTGTACGAGGACACCTGACCTGTTCCTTCTCCGGGTATACCTTCAAAGCAC
Db	2104	TCGATGGAAAAACCCAGGCTTATGGAATTTGGGTGCCCACAACTCAGACTTTCGGAACAGA	2163	TCGATGGAAAAACCCAGGCTTATGGAATTTGGGTGCCCACAACTCAGACTTTCGGAACAGA
Qy	2124	TCTATGGAGAACCCCGGCTGTGGATTCTTGGGCTGCCACAACAGCGACTTCGCGAACCGC	2183	TCTATGGAGAACCCCGGCTGTGGATTCTTGGGCTGCCACAACAGCGACTTCGCGAACCGC
Db	2164	GGCATGACCGCCTTACTGAAGTTTCTAGTTGTGACAAGAACAACACTGGTGTATTACGAG	2223	GGCATGACCGCCTTACTGAAGTTTCTAGTTGTGACAAGAACAACACTGGTGTATTACGAG
Qy	2184	GGCATGACTGCCCTGCTGAAGTCTCCAGCTGGACACAAGACCCGGCGACTACTACGAG	2243	GGCATGACTGCCCTGCTGAAGTCTCCAGCTGGACACAAGACCCGGCGACTACTACGAG
Db	2224	GACAGTTATGAAGATATTTTCAGCATACTTGTCTGAGTAAAAACAATGTCATTGAACCAAGA	2283	GACAGTTATGAAGATATTTTCAGCATACTTGTCTGAGTAAAAACAATGTCATTGAACCAAGA
Qy	2244	GACAGCTACGAGGACATCTCCGCCCTACTGCTGTCCAAGAACACGCCATTCGAGCCCCGC	2303	GACAGCTACGAGGACATCTCCGCCCTACTGCTGTCCAAGAACACGCCATTCGAGCCCCGC
Db	2284	AGCTTCTCCAGAAATTCAGACACACCTTAGCACTAGGCAAAAGCAATTTAATGCCACC	2340	AGCTTCTCCAGAAATTCAGACACACCTTAGCACTAGGCAAAAGCAATTTAATGCCACC
Qy	2304	TCCTTCTCCCAAACTCCCGCCACCCAGCACCGGCTCAGAAAGCAGTTCAACGCCACC	2360	TCCTTCTCCCAAACTCCCGCCACCCAGCACCGGCTCAGAAAGCAGTTCAACGCCACC

RESULT		7	E00527	8831 bp	RNA	PAT	26-NOV-1996
LOCUS			Complete cDNA encoding human factor VIII.				
DEFINITION			E00527				
ACCESSION			g2168806				
NID			JP 1985243023-A/2.				
KEYWORDS			Homo sapiens.				
SOURCE			Homo sapiens				
ORGANISM			Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;				


```
REFERENCE      1 (bases 1 to 8831)
AUTHORS        Keipon,D.J., Roon,R.M., Biihaa,G.A. and Utsudo,U.A. .
TITLE          MANUFACTURE OF FUNCTIONAL VIII FACTOR
JOURNAL        Patent: JP 1985243023-A 2 03-DEC-1985;
                GENENTECH INC

COMMENT
OS      Homo sapiens
PN      JP 1985243023-A/2
PD      03-DEC-1985
PF      19-APR-1985 JP 1985085295
PR      20-APR-1984 US 84    602312
PI      DANIERU JIEFURII KEIPON, RICHIIYAADO MAAKU ROON, PI    GOODON
AREN     BIIHAA,
PI      UIRIAMU AAUIN UTSUDO
PC      A61K37/04,A61K35/12,A61K35/74,C07H21/04,C07K13/00,C07K15/04,
PC      C12N5/00,
PC      C12N15/00//C12P21/00,(C12N5/00,C12R1:91);
CC      strandedness: Single;
CC      topology: Linear;
CC      hypothetical: No;
CC      anti-sense: No;
CC      *source: cell_line=CH-2;
CC      *source: chromosome/segment=X-chromosome;
FH      Key      Location/Qualifiers
FH
FH      5'UTR      1..109
FH      sig_peptide 110..166
FH      /product='factor VIII signal peptide' FT
FH      mat_peptide 167..7262
FH      /product='factor VIII'
FH      CDS      110..7265
FH      /product='factor VIII'
FH      3'UTR      7266..<8967
FH      polyA_signal 8948..8953
FH      polyA_site 8968.
FEATURES
source      Location/Qualifiers
            1..8831
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  2807 a 1865 c 1800 g 2359 t
ORIGIN

Query Match      24.4%; Score 1085; DB 18; Length 8831;
Best Local Similarity 73.8%; Pred. No. 0.00e+00;
Matches 1681; Conservative 0; Mismatches 596; Indels 0; Gaps 0;

Db  167 GCCACCAGAAGATACTACCTGGGTGCAGTGGAACTGTCATGGGACTATATGCAAGTGAT 226
      ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  84 GCCACCCGCCGTACTACCTGGGCGCGTGGAGCTGTCTGGGACTACATGCAGACGCAC 143

Db  227 CTCGGTGAGCTGCGTGTGGACGCAAGATTTCCTCCTAGAGTGCCAAATCTTTTCCATTC 286
      || || ||||| || ||||| || || || || || || || || || || || || || ||
Qy  144 CTGGGGAGACTCCCGTGGACGCCGGCTTCCCGCCCGCGTGGCCCAAGAGCTTCCCGCTTC 203

Db  287 AACACCTCAGTCGTGTACAAAAAGACTCTGTTTGTAGAATTACGGGATCACCTTTTCAAC 346
      ||||| || ||||| || || ||||| || || ||||| || || ||||| || || |||||
Qy  204 AACACCACGGTGTGTACAAGAAACCCTGTCTGGAGTTTCCCGCCCGCACCACCTGTTCAAC 263

Db  347 ATCGCTAAGCCAGGCCACCCCTGGATGGGTCTGCTAGGTCCTACCATCCAGGCTGAGGTT 406
      || || ||||| || || ||||| ||||| || || ||||| || || ||||| |||||
Qy  264 ATTGCCAAGCGCGCCCCCCTGGATGGGCTGTCTGGGCCCCCACCATTCCAGGCCAGGTG 323

Db  407 TATGATACAGTGTTCATTACACTTAAGAACATGGCTTCCCATCCTGTGAGTCTTCATGCT 466
      || || || ||||| || || || ||||| ||||| || || ||||| || || || || ||
```

```
Qy  324 TACGACACCGTGGTGATCACCCCTGAAGAACATGCCCCAGCCACCCCGTCAGCGCTGCAGGCC 383

Db  467 GTTGTGTATCCTACTGGAAGCTTCTCAGGGAGCTGAATATATGATCAGACCAGTCAA 526
      || || || ||||| ||||| || || || || || || || || || || || || || ||
Qy  384 GTGGCGTGAGCTACTTGGAAAGCCAGCGGGCCCGAGTACGAGCACACACCTGCCAG 443

Db  527 AGGAGAAAGAGATGATAAAGTCTTCCTTGGTGGAAAGCATACATATGCTGGCAGGTC 586
      | ||||| || || || || ||||| || || ||||| || || ||||| || || ||||| ||
Qy  444 CCCGAGAAGGAGCAGCACAAAGTGTTCCCGGGGGGAGCCACACCTACGTGTGGCAGGTG 503

Db  587 CTGAAAGAGAAATGGTCCAATGGCCTCTGACCCACTGTGCTTACCTACTCATATCTTTCT 646
      || || ||||| || || ||||| ||||| ||||| ||||| ||||| || || || || ||
Qy  504 CTTAAGGAGAAGCGGCCCTATGGCCAGCCACCCCTGTGCTTGACCTACAGCTACCTGAGC 563

Db  647 CATGTGCACCTGGTAAAGACTTGAATTCAGGCTCATTTGGAGCCCTACTAGTATGTAGA 706
      || ||||| ||||| || || ||||| || || || || || || || || || || || || ||
Qy  564 CACGTGCACCTGGTGAAGGATCTGAACAGCGGGGTGATCGGCGCCCTGCTGGTGTGCGC 623

Db  707 GAAGGAGTCTGGCCCAAGGAAAGACACAGACCTTGCACAAATTTATACTACTTTTGCT 766
      || || || ||||| ||||| || || ||||| ||||| || || || || || || || ||
Qy  624 GAGGCGCCTGGCCAAAGGAGAAACCAGACCCCTGCACAAGTTTCATCCTGCTGTGCCC 683

Db  767 GTATTTGATGAAGGGGAAAAGTTGGCACTCAGAAACAAAGAACTCCTTGATGCAGGATAGG 826
      || || || || ||||| || ||||| || || ||||| || ||||| || ||||| || ||
Qy  684 GTGTTCCAGGAGGGAAGAGCTGGCACAGCGAGACTAAGAACAGCCTGATGCAGGACCGC 743

Db  827 GATGTGCATCTGCTCGGCGCTGGCCTAAATTCACACAGTCAATGGTTATGTAAACAGG 886
      || || || || || ||||| || ||||| || || ||||| || || || || || || || ||
Qy  744 GACGCGCCAGCGCCCGCCTGGCCCAAGATGCACACCGTTAACGGCTAGCTGAACCGC 803

Db  887 TCTTGCCAGGTCTGATTGGATGCCACAGGAATCAGTCTATTGGCATGTGATTGGAATG 946
      ||||| || ||||| || ||||| || || || ||||| || || ||||| || || || || ||
Qy  804 AGCTGCGCGGCTGATCGGCTGCCACCGCAAGAGCGTCTACTGGCAGCTCATCGGCGATG 863

Db  947 GGCACCACTCCTGAAAGTGCACTCAATATTCTCGAAGGTCACACATTTCTTGTGAGGAAC 1006
      ||||| ||||| ||||| || ||||| || || ||||| || || ||||| || || || || ||
Qy  864 GGCACCACCCTGAGGTGCACAGCATCTTCTCGAGGGGCACACCTTCTTGGTGGCAAC 923

Db  1007 CATCGCCAGGCTCCTTGGAAATCTCGCCAAATTACTTCTTACTGCTCAACACTCTTTG 1066
      || ||||| || || || || || || || || || || || || || || || || || || ||
Qy  924 CACGCGCAGGCAGCCTGGAGATCAGCCCCATCACCCTTCTGACTGCCAGACCCCTGCTG 983

Db  1067 ATGGACCTTGCACAGTTTCTACTGTTTGTGTCATATCTTCCCACCAACATGATGGCATG 1126
      ||||| || ||||| || ||||| || || || || || || || || || || || || || ||
Qy  984 ATGGACCTAGGCCAGTTCCTGCTGTTCTGCCACATCAGACGCGCACCGACGCGGCGATG 1043

Db  1127 GAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCAACTACGAATGAAAAATAAT 1186
      || ||||| || || || ||||| || ||||| ||||| || || ||||| || || || || ||
Qy  1044 GAGGCTTACGTGAAGGTGGACAGCTGCCCGAGGAGCGCCAGCTGGGCATGAAGAACAAC 1103

Db  1187 GAAGAGCGGAAGACTATGATGATGATCTTACTGATTCTGAAATGGATGTGGTCAGGTTT 1246
      || || || || ||||| || || || || || || || || ||||| || || || || ||
Qy  1104 GAGGAGCGGAGGACTACGACGACGACTGACCCACAGCGAGATGGATGTGCTACGCTTC 1163

Db  1247 GATGATGACAACCTCTCCTTCTTTATTCAAAATTCGCTCAGTTGCCAAGAGCATCTTAA 1306
      || || ||||| || || ||||| || || ||||| || || || ||||| ||||| |||||
Qy  1164 GACGACCAACAAGCCCCAGCTTCATCCAGATCCGACGCTGGCCAAAGACACCCCTAAG 1223

Db  1307 ACTTGGGTACATTACATTGCTGCTGAACAGGAGGAGCTGGGACTATGCTCCCTTAGTCCTC 1366
      || ||||| || ||||| || || || ||||| ||||| || || ||||| || || || || ||
Qy  1224 ACCTGGGTGACTACATCGCGCCGAGGAGGAGGAGCTGGGACTACGCCCGCTAGTACTG 1283

Db  1367 GCCCCCGATGACAGAACTTATAAAAGTCAATATTGAACAATGGCCCTCAGCGGATGGT 1426
      ||||| || || || || || || || || || || || || || || || || || || || ||
```


/note="preprocoagulation factor VIII:C"
/codon_start=1
/db_xref="GDB:C00-119-124"
/db_xref="PID:g182818"
/translation="MQIELSTCFLLRRCFSATRRYYILGAVELSWDMQSDLIGELP
VDARFPPRPKSPFPNTSVYKKTLEVEFDHLENIAKPRPPWMLLGPITQAEVYDT
VVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSOREKEDDKVPPGGSHYYWQVL
KENGPMASDPLCLTYSYLSHVDLVKDLNLSGLICALLVCREGSLAKEKTQTLHKFLLF
AVFDEGKSHSETKNSLMQDRDAASARAWPKMHTVNGVNRSLPGLICGRKSVYMHV
IGMGTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLMLDLGQFLFCHISSH
QHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSMDVVRDNDSPSFIQRS
VAKKHPKTWVHYIAAEEDWDYAPIVLAPDDRSYKSYLNNGPQIRGRKYKVRWAY
TDTEFKTREAIQHESGILGPLLYGEVDTLLIIFKNQASRPYNIYPHGITDVRPLYSR
RLPKGVKHLKDFPILPGEIEFYKKTWTVVEDGPTKSDPRCLTRYYSFVNMERDLASGL
IGPILLICYKESVDQRGNQIMSDKRNVIILFVFDENRSHYLTENIQRF LNPAGVQLED
PEEQASNIMHSINGVVFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKV
YEDTLTLFPFSGETVFMSEMPGLWILGCHNSDFRNRMGTALLKVVSSCDKNTGYED
SYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTIPENDIEKTDPFVFAHRTPM
PKIQNVSSDILLMLLRQSPPHGLSISDLQEAKEYTFSDDPSPGAIDSNNSLSEMTFH
RPQLHSGDMVFTPESGLPLRNEKLGTAAELKCLDFKVPSSYNNLSTIPSDNLA
ACTDNTSSLGPPSMPVHYDSQLDITLFGKKSPLTESGPLSISEENNDSKLLSEGLM
NQESSWGKNVSTESGRLEFKRAHPALLTKDNALFKVSIISLTKNTSNNSATNR
KTHIDGPSLLIENSPSVWQNILESDFEKKVTPLIHRLMLDMKNATALRLNHSNKTT
SSKNMEMVQKKKEGIPPPDAQNPDMSFFKMLFPE SARHIQRTHGKNSLNSGQSPK
QLVSLGPEKSVGEQNF LSEKNKVVGKGEFTKDVGLKEMVFPSSRNLF LTNLDNLHEN
NTHNQEKKIQEEIEKKETLIQENVVLPQIHVTGTGNFMKNFLILSTRQNVESYDGA
YAPVLQDFRS LNDSTNRTKHTAHFSKKGEEENLEGLNQTKQIVEKYACTTRISPNT
SQQNFVQRSKRALKQFLRPLEETELEKRIIVDDTSTQWSKNMKHLTPSLTQIDYNE
KEKGAIQTQSP LSDCLTRSHSIPQANRSLPIAKVSSFPPIRP IYILTRVLFPQDNSHLP
AASRYKSDSGVESHSFLQCAKKNLISLAILTEMTDQREVSGLSGTSATNSVTYKKV
ENTVLPKPDLPKTSCKVELLPKVHIYQKDLFFETETNSGPHLDLVEGSLLOGTEGAI
KNEANRPCKVPFLRVATESAKTPSKLLDP LADWNHYQTQIPKEEMKSEKSPKTA
FKKDOTILSINACESNHAIAINEGQNKPEIEVTWAKQGRTERLCSQNPPVILKRQRE
IYRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLW
DYGMSSSHVLRNRAQSGVPQFKKVVQEFTDGSFTQPLYRGELNEHLGLLGPYIRA
EVEDNIMVTFRNQASRPYSFYSSLSIYEEDQROGAEPKRFVKPNETKTYFWKVQHMM
APTQDEFCKAWAYFSDVDLEKDVHSLIGPLILVCHTNTLNPAGHRQVTVQEFALFFT
IFDETQSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQR
IRWYLLSMGSNENIHSIHFSCHVFTVRKKEEYKMALYNLYPGVETVEMLP SKAGIWR
VECLIGEHLHAGMSTFLVYSNKKQTP LGMASGHIRDFOITASGOYQGWAPKLARIHY
SGSINAWSTKEPFSWIKVDLLAPMIHGIKTQCARQKFSLSYISQFIIMYSLDGRKKWQ
TVRGNSTGTLMVFFGNVDSSGIKHNIENPPIIARIYIRLPHTHYSIRSTLRMELMGCDL
NSCSMP LGMESKAISDAQITASSYFTWFAFTWSPSKARLHLHQGRSNARPPQVNNPKEM
LQVDFQKTMKVTVGTTQGVKSLLTSMVYKEFLISSQDGHQWTLFFQNGKVKVFQGNQ
DSFTPVVNSLDPP LLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY"
898. .7224
/gene="F8C"
/note="coagulation factor VIII:C"
1042
/gene="F8C"
/note="g in wt; a in hemophiliac [2]"
1043
/gene="F8C"
/note="a in wt; g in hemophiliac [2]"
6853
/gene="F8C"
/note="c in wt; t in hemophiliac [3]"
BASE COUNT 2860 a 1910 c 1848 g 2411 t
ORIGIN 185 bp upstream of SacI site.

Query Match 24.4%; Score 1085; DB 21; Length 9029;
Best Local Similarity 73.8%; Pred. No. 0.00e+00;
Matches 1681; Conservative 0; Mismatches 596; Indels 0; Gaps 0;

Db 229 GCCACCACAAGATACTACTGGGTGCAGTGGAACTGCTCATGGCACTATATGCAAAAGTGAT 288
||||| | | ||||||||| | | ||||| ||||| ||||||| ||||| | |
Qy 84 GCCACCCCGCTACTACTGGCGCGGTGGAGCTGTCTCTGGGACTACATGCAGACGCAC 143

Db 289 CTCGGTGAGCTGCCTGTGACAGCGCAAGATTTCCTCCTAGAGTGCCAAAATCTTTTCCATT 348
| | | ||||| | | ||||||| | | | | | | | | | ||||| | |
Qy 144 CTGGCGAGCTCCCCGTGGACGCCCGCTTCCCCCCCCCGGTGCCCAAGAGCTTCCCCCTTC 203

Db 349 AACACCTCAGTCTGTACAAAAAGACACTCTGTCTTGAGAAITTCACGGATCACCTTTTCAAC 408
||||| | | ||||||| | | | | | | | | | ||||| | | |||||
Qy 204 AACACCAGCTGGTGTACAAGAAAAACCTGTTCGTGGAGTTACCGGACCACCTGTTCAAC 263

Db 409 ATCGCTAAGCCAGGCCACCCCTGGATGGGTCTGTGTAGGTCCTACCATCCAGGCTGAGGTT 468
| | | ||||| | | ||||||||| | | | | | | | | | |||||
Qy 264 ATTGCCAAGCCGCGCCCCCTGGATGGGCTGTCTGGGCCCCACCATCCAGGCGGAGGTG 323

Db 469 TATGATACAGTGTCTATTACACTTAAGAACATGGCTTCCCATCTCTGCTCAGTCTTTCATGCT 528
| | | ||||| | | | | | ||||||||| | | | | | | | | | |||||
Qy 324 TACGACACCGTGTGTATCACCTGAAGAACATGGCCAGCCACCCCGTCAGCCTGCACGCC 383

Db 529 GTTGTGTATCTCTACTGGAAGCTTCTGAGGGAGCTGAATATGATGATCAGACCAAGTCAA 588
| | | | | ||||||| | | ||||| | | | | | | | | | ||||| | |
Qy 384 GTGGGCTGAGCTAGTCTGGAAGCCACGAGGGGCGCCAGTACGACGACGACGCTCCGAG 443

Db 589 AGGAGAAAGACATGATAAAGTCTTCCCTGGTGAAGCCATACATATGTCTGGCAGGTC 648
| ||||| | | | | | | | | | ||||| | | | | | | | | | |||||
Qy 444 CCGAGAAAGGACGACGACAAGGTGTTCCCGGGGGGAGGCCACACCTACGTGTGGCAGGTG 503

Db 649 CTGAAACAGAAATGGTCCAAATGGCTCTGACCCACGTGTGCCTTACCTACTCATATCTTTCT 708
| | | ||||| | | | | | ||||| ||||| ||||||| ||||| | |
Qy 504 CTTAAGCAGAACGGCCCTATGGCCAGCAGACCCCTGTGCCTGACCTACAGCTACCTGAGC 563

Db 709 CATGTGACCTGTGTTAAAGACTTGAATTCAGGCCCTCATTTGAGGCCCTACTACTATGTAGA 768
| | ||||||||| | | | | | | | | | | | | | | | ||||| | |
Qy 564 CACGTGACCTGGTGAAGGATCTGAACAGCGGGCTGATCGGCGCCCTGCTGGTGTGTCGC 623

Db 769 GAAGGGAGTCTGCCAAGGAAAAGACACAGACCTTGCACAAAATTTATACTACTTTTTGCT 828
| | | | | ||||||||| | | | | | ||||| | | ||||| | | | | |
Qy 624 GAGGGACCTGGCCAAGGAGAAAAACCCAGACCCCTGCACAAGTTTCATCTGCTGTTCGCC 683

Db 829 CTATTTTATGAAGGGAAAAGTTGGCACTCAGAAAACAAAGAACTCCTTGATGCAGGATAGG 888
| | | | | ||||| | | ||||| | | | | | | | | | ||||| | |
Qy 684 GTGTTCCAGGAGGGAGAGCTGGCACAGCGAGACTAAGAACAGCCTGATGCAGGACCGC 743

Db 889 GATGCTGCATCTGCTCGGCGCTGGCCTAAAATGCACACAGTCAATGGTTATGTAACAGG 948
| | | | | ||||| | | ||||||| | | ||||||| | | | | | | | |
Qy 744 GACGCCGCCAGCGCCCGCCCTGGCCCAAGATGCACACCGTTAACGGCTACGTGAACCGC 803

Db 949 TCTCTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATG 1008
||||| | | ||||| | | ||||||| | | ||||| | | ||||| | |
Qy 804 AGCCTGCCCGGCTGATCGGCTGCCACCCGCAAGCGGTGACTGGCACGTCATCGGCATG 863

Db 1009 GGCACCACTCCTGAAGTGCACCTCAATATTTCCTGAAGGTGCACACATTTCTTCTGAGGAAC 1068
||||||| ||||| ||||| | | ||||| | | | | | | | | | |||||
Qy 864 GGCACCAACCCCTGAGGTGCACAGCATCTTCCTGGAGGGGCCACACCTTCCTGGTGGCGAAC 923

Db 1069 CATGCCAGGCGCTTGGAAATCTCGCCAATAACTTTCCTTACTGCTCAAAACACTCTTG 1128
| | ||||||| | | ||||| | | | | | | | | | ||||| | | | | |
Qy 924 CACCGCCAGGCCAGCCTGGAGATCAGCCCCATCACCTTCCTGACTGCCCCAGACCCCTGCTG 983

Db 1129 ATGGACCTTGGACAGTTTCTACTGTCTTTTCTCATATCTCTTCCCAACCAACATGATGGCATG 1188
||||||| | | ||||| | | ||||| | | | | | | | | | ||||| | |
Qy 984 ATGGACCTAGGCCAGTTCTCTGTGTCTTGCCACATCAGCAGGCCACCGACGACGCGCATG 1043

Db 298 TATGATACAGTGTCTCATTTACACTTAAGAACATGGCTTCCCATCTGTCCATCTTCTGATCTTCATGCT 357
Qy 324 TACGACACCGTGTGTGATCACCCCTGAAGAACATGGCCAGCCACCCCGTAGCCTTGACGCC 383

Db 358 GTTGGTGTATCTCTGAAAGCTTCTGAGGAGCTGAATATGATGATCAGACAGTCAA 417
Qy 384 GTGGCGTGAAGTCTGGAAGGCCAGCGAGGGCGCCGAGTAGCAGCAGCAGCTCCAG 443

Db 418 AGGAGAGAAAGATGATFAAAGTCTTCCCTGGTGAAGCCATACATATCTCTGGCAGGTC 477
Qy 444 CGCGAGAGGAGCAGCAGCAGAGGTGTTCCCGGGGGGAGCCACACCTACCTGTGGCAGGTG 503

Db 478 CTGAAAGAGAAATGGTCCCAATGGCTCTGACCCACTGTGCCCTTACCTACTCATATCTTTCT 537
Qy 504 CTTAAGCAGAACGGCCCTATGGCCAGCGACCCCTGTGCCCTGACCTACAGCTACCTGAGC 563

Db 538 CATGTGGACCTGCTAAAGACTTGAATTCAGGCCCTCATTTGGAGCCCTACTAGTATGTAGA 597
Qy 564 CACGTGACCTGCTGAAGCATCTGAACAGCGGGCTGATCGGCGCCCTGCTGGTGTGTCGC 623

Db 598 GAAAGGAGTCTGCCCAAGGAAAGACACAGACCTTGCACAAATTTATACTACTTTTGTCT 657
Qy 624 GAGGCGACCTGGCCCAAGCAGAAACCCAGACCCCTGCACAAAGTTCACTCTGCTGTTCGCC 683

Db 658 GTATTTGATGAAGGGAAAGTTGGCACTCAGAAACAAAGAACTCCCTTGATGCAGATAGG 717
Qy 684 GTGTTCCAGCAGGGGAAGACTGGCACAGCCAGACTAAGAACAGCCCTGATGCAGGACCCG 743

Db 718 GATGCTGCATCTGCTCGGGCTGGCTTAAATGCACAGTCAATGGTTATGTAAACAGG 777
Qy 744 GACGCCGCCAGGCCCGCCCTGGCCCAAGATGCACACCCGTTAACGGCTACGTGAACCCG 803

Db 778 TCTCTGCAGGTCTGATTGATGCCACAGGAAATCACTCTATTGGCATGTGATTGGAATG 837
Qy 804 AGCCTGCCCCGCTGATCGGCTGCCACCGCAAGCGGTGTACTGGCACGTCTCGGCATG 863

Db 838 GGCACCACTCTCTGAAAGTGCACCTCAATATTCTCGAAGGTCAACATTTCTTGTGAGGAAC 897
Qy 864 GGCACCAACCCCTCAGGTGCACAGCATCTTCTCGAGGGCCACACCTTCTCTGGTGGCAAC 923

Db 898 CATCGCCAGGCTCTTGGAAATCTGCCCAATAACTTCTTACTGCTCAAAACACTCTTG 957
Qy 924 CACCGCCAGGCCAGCCTGGAGATCAGCCCCATCACCTTCTTGACTGCCAGACCCCTGCTG 983

Db 958 ATGGACCTTGGACAGTTTCTACTGTTTTGTTCATATCTCTTCCACCAACATGATGGCATG 1017
Qy 984 ATGGACCTAGGCCAGTTCTCTGCTTCTGCCACATCAGCAGCCACCAGCAGCGGCATG 1043

Db 1018 GAAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCAACTACGAATGAAATAAT 1077
Qy 1044 GAGGCTTACGTGAAGGTGGACAGCTGCCCGGAGGAGCCCCAGCTGGCGATGAAGAACAA 1103

Db 1078 GAAAGACGGGAAGACTATGATGATCTTACTGATTTCTGAAATGGATGGTGGTCAAGTTT 1137
Qy 1104 GAGGAGCCGAGACTACGACGACACCTGACCGCAGCGGAGATGGATGCTCGTAGCGCTTC 1163

Db 1138 GATGATGACAACTCTCCTTCTTATCCAAATTCGCTCAGTTGCCAAGAGCATCTCTAAA 1197
Qy 1164 GACGACCAACAGCCCCAGCTTCATCCAGATCCGCGAGCGTGGCCCAAGAGCAACCTAAG 1223

Db 1198 ACTTGGTACATTAATTGCTGCTGAAGAGGAGCTGGGACTPATGCTCCCTTAGTCTC 1257
Qy 1224 ACCTGGTGCATACATCGCCCGCGAGGAGGAGCTGGGACTACGCCCGCTAGTACTG 1283

Db 1258 GCCCCCGATGACAGAAAGTTATAAAAGTCAATATTGAAACAAATGGCCCTCAGCGGATTTGTT 1317
Qy 1284 GCCCCCGACGACCGGAGCTACAGAGCCAGTACTTGAACAACGGCCCCCAGCGCATCGGC 1343

Db 1318 AGGAAGTACAAAAAAGTCCGATTTATGGCATACACAGATGAAAACCTTTAAGACTCGTGAA 1377
Qy 1344 CGCAAGTACAGAAGGTGCGCTTCATGGCCTACACCGCAGAGACTTTCAGACCCCGCGAG 1403

Db 1378 GCTATTCAGCATGAATCAGGAATCTTTGGGACCCTTTACTTTATGGGGAAGTTGGAGACACA 1437
Qy 1404 GCCATCCAGCAGAGTCCGGCATCTCTGGCCCCCTGCTGTACGGCAGGTTGGCGCAGACC 1463

Db 1438 CTGTTGATTATATTTAAGAATCAAGCAAGCAGACCATATAACATCTTACCCCTCACGGAATC 1497
Qy 1464 CTGCTGATCATCTTCAAGAACCGAGCCAGCGCCTACAACATCTACCCCCACGGCATC 1523

Db 1498 ACTGATCTCGTCTCTTTCTATTCAAGGAGATTACCAAAAGGTGTAAAACATTTGAAGGAT 1557
Qy 1524 ACCGACGTGGCCCCCTGTACAGCCGCCCTGCCAAGGGCGTGAAGCACCTGAAGGAC 1583

Db 1558 TTTCCAAATCTGCCAGGAGAAATATTCAAATATAAATGGACAGTACTGTAGAAGATGGG 1617
Qy 1584 TTCCCATCTTCCCCGGCGGAGATCTTCAAGTACAACTGACCCGTGACCCGTGGAGGACGGC 1643

Db 1618 CCAACTAAATCAGATCTCTGGTGCCTGACCCGCTATTACTCTAGTTTCGTTAATAATGGAG 1677
Qy 1644 CCCACCAAGAGCGACCCCGCTGCCCTGACCCGCTACTACAGCAGCTTCGTGAACATGGAG 1703

Db 1678 AGAGATCTAGCTTCAGGACTCAATTGGCCCTCTCTCTCATCTGCTACAAAGAACTGTAGAT 1737
Qy 1704 CGCGACCTGGCCTCCGGACTGATCGGCCCCCTGCTGATCTGCTACAAGGAGAGCGTGGAC 1763

Db 1738 CAAAGAGGAAACAGATAAATGTCAGACAAAGAGGAATGTCATCTCTGTTTCTGTATTGAT 1797
Qy 1764 CAGCGGGCAACCAATCATGAGCGGACAAGCGCAACGTGATCTGTTCAAGCGTGTTCGAC 1823

Db 1798 GAGAACCGAAGCTGGTACTCACAGAGAAATATACAACGGCTTCTCCCCAAATCCAGCTGSA 1857
Qy 1824 GAGAACCGCAGCTGGTATCTGACCGGAGAAACATCCAGCGCTTCTTGCCCAACCCCCGCTGGC 1883

Db 1858 GTGCAGCTTAGGATCCAGAGTTCCAAGCCTCCAACATCATGCACAGCATCAATGGCTAT 1917
Qy 1884 GTGCAGCTGGAAGATCCCGAGTTCAGGGCCAGCAACATCATGCAAGCATCAACGGCTAC 1943

Db 1918 GTTTTTCATAGTTTGCAAGTTGTGAGTTTGTGTTTTCGATGAGGTGGCATPACTGGTACATTTCTA 1977
Qy 1944 GTGTTCCACAGCCTGCAGCTGAGCGTGTGCCTGCATGAGGTGGCTTACTGGTACATCCTG 2003

Db 1978 AGCATTTGGAGCACAGACTGACTTCTCTTTCTGTCTTCTTCTTGATATACCTTCAAACAC 2037
Qy 2004 AGCATCGCGGCCAGACCGGACTTCTTGAGCGTGTCTTCTCCGGGTATACCTTCAAGCAC 2063

Db 2038 AAAATGCTCTATGAAGACACACTCACCCCTATTCCCATTTCTCAGGAGAAAACCTGCTTTCATG 2097
Qy 2064 AAGATGCTGTACGAGGACACCCCTGACCCCTGTTCCCTTCTCCGGCGAGACTGTGTTTCATG 2123

Db 2098 TCGATGGAACCCAGGTCTATGGATTCTGGGGTGCCACAACCTCAGACTTTCGGAACAGA 2157
Qy 2124 TCTATGGAGAACCCCGCCTGTGGATTTCTGGGCTGCCACAACAGCGACTTCCGCAACCCGC 2183

Db 2158 GGCATGACCGCCTTACTGAAGGTTTCTAGTTGTGACAAGAACACTGGTGTATTATTACGAG 2217
Qy 2184 GGCATGACTGCCCTGCTGAAAAGTCTCGAGCTGCCACAAGAACACCGGCGGACTACTACGAG 2243

Db	2218	GACAGTTATCAAGATATTTTCAGCATACTTGCCTGAGTAAAAACAATGCCATTGAACCAAGA	2277
Qy	2244	GACAGCTACGAGGACATCTCCGCCCTACCTGCTGTCTCAAGAACAAACGCCATCGAGCCCCGC	2303
Db	2278	AGTTTCTCCAGAAATTCAGACACCCCTAGCACTAGGCAAAAGCAAATTTAATGCCACC	2334
Qy	2304	TCCTTCTCCAAACTCCCGCCACCCAGACGCGCTCAGAAGCAGTTCAACGCCACC	2360
RESULT	10		
LOCUS	I08345	7440 bp	PAT 14-NOV-1994
DEFINITION	Sequence 2 from Patent WO 8501961.		
ACCESSION	I08345		
NID	g588944		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 7440)		
TITLE	Toole,J.J.J.		
JOURNAL	PRODUCTION OF FACTOR VIII AND RELATED PRODUCTS		
FEATURES	Patent: WO 8501961-A 2 09-MAY-1985;		
source	Location/Qualifiers		
	1. .7440		
	/organism="unknown"		
BASE COUNT	2316 a	1631 c	1554 g 1936 t 3 others
ORIGIN			
Query Match 24.3%; Score 1081; DB 18; Length 7440;			
Best Local Similarity 73.7%; Pred. No. 0.00e+00;			
Matches 1679; Conservative 0; Mismatches 598; Indels 0; Gaps 0;			
Db	216	GCCACCAGAAGATACTACCTGGGTGCAGTGCAACTCTCATGGGACTATATGCAAAGTCAT	275
Qy	84	GCCACCCGCCGTACTACCTGGGCGCGTGGAGCTGTCTGGGACTACATGCAGAGCGAC	143
Db	276	CTCGGTGAGTGCCTGTGGACGCAAGATTTCTCTCCTAGAGTGC AAAATCTTTTCCATTC	335
Qy	144	CTGGCGAGCTCCCGTGGACGCCCGCTTCCCCCGCGTGCCCAAGAGCTTCCCCTTC	203
Db	336	AACACCTCAGTCTGTACAAAAAGACTCTGTTGTAGAATTCACGGTTCACCTTTTCAAC	395
Qy	204	AACACGAGGTGTTGTACAAGAAACCCTGTTCTGTGAGTTTCCCGACCACTGTTTCAAC	263
Db	396	ATCGCTAAGCCAGGCCACCCTGGATGGGTCTGCTAGGTCTCTACCATCCAGGCTCAGGTT	455
Qy	264	ATTGCCAAGCCGCGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAGGCCGAGTG	323
Db	456	TATGATACAGTGGTCATTACACTTAAGAACATGGCTTCCCATCCTGTCACTTTCATGCT	515
Qy	324	TACGACACCGTGGTGATCACCCCTGAAGACATGGCCAGCCACCCCGTCAGCCTGCACGCC	383
Db	516	GTGGGTATCTCTACTGGAAGCTTCTGAGGAGCTGAATATGATATCAGACCACTCAA	575
Qy	384	GTGGCGGTGAGCTACTGGAAGCCAGCGAGGCGCGCAGTACGACGACAGACCTCCAC	443
Db	576	AGGAGAGAAAGATGATATAAGTCTTCCCTGGTGAAGCCATACATATGCTGGCAGGTC	635
Qy	444	CGGAGAGGAGGACGACAGGTGTTCCCGGGGGAGCCACACCTACGTGTGGCAGGTG	503
Db	636	CTGAAAGAGAAATGGTCCAAATGGCCCTCTGACCCACTGTGCGCTTACCTACTCATCTTCT	695
Qy	504	CTTAAGGAGAACGGCCCTATGGCCACCGACCCCTGTGCTGACCTACAGCTACCTGAGC	563

Db	696	CATGTGCACCTGCTAAAACACTTCAATTACGCCCTCATTTGAGAGCCCTACTAGTATGTACA	755
Qy	564	CACGTGCACCTGGTGAAGGATCTGAACACAGGGGCTGATCGCGCCCTGCTGCTGTGTGCG	623
Db	756	GAAGGGAGTCTGCCCCAAGSAAACACACACACACCTTCGCACAAATTTATACTACTTTTGT	815
Qy	624	GAGGGACGCTGGCCAAAGGAGAAAACCAGACCCCTGCACAAGTTCAFCCTGCTGTTCGCC	683
Db	816	GTATTTGATGAAGGGAAAAGTTGSCACTCAGAAAAACAAAGAACTCCTTTGATGCAGGATAG	875
Qy	684	GTGTTCCAGCGAGGGGAACAGCTGGCACAGCGAGACTAAGAACAGCCTGATGCAGGACCG	743
Db	876	GATGCTGCATCTGCTCGGGCCTGGCCTAAATATTCCTCGAAGGTCACACATTTCTTCTGAGGA	935
Qy	744	GACGCCCCAGCGCCCCGCGCTGGCCCCAAGATGCACACCCGTTAACGGCTACGTGAACCG	803
Db	936	TCTCTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTTGGGATGTGATTGGAATG	995
Qy	804	AGCCTGCCCGGCTGATCGGCTGCCACCGCAAGACCGTGTACTGGCACGTATCGGCATG	863
Db	996	GGCACCACTCCTGAAGTGCACCTCAATATTCCTCGAAGGTCACACATTTCTTCTGAGGAAC	1055
Qy	864	GGCACCCCTGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTCTGTCGCGCAAC	923
Db	1056	CATGCCCAGGCTCCTTGGAAATCTCGCCAATAACTTTCTTACTGCTCAACACACTCTTG	1115
Qy	924	CACCGCCAGGCCCTGGAGATCAGCCCCATCCTTCTCTGACTGCCCCAGACCCCTGCTG	983
Db	1116	ATGGACCTTGGACAGTTTCTACTGTTTTTGTCTATATCTCTTCCCACCAACATGATGGCAT	1175
Qy	984	ATGGACCTAGGCCAGTTCTCTGTTCTGCCCATCAGCAGCCACCAGCAGCGGCATG	1043
Db	1176	GAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCAACTACGAATGAAAAATAAT	1235
Qy	1044	GAGGCTTACGTGAAGGTGGACAGCTGCCCGGAGGAGCCCGAGCTGCGCATGAAGAACAA	1103
Db	1236	GAAGAAGCGGAAGACTATGATGATCTTACTGATTTCTGAAATGCGATGTGTCAGGTTT	1295
Qy	1104	GAGGAGCGGAGGACTAGCAGCAGGACCTTGCCACACAGCAGATGCGTGTCTACGCTTC	1163
Db	1296	GATGATCACAACCTCTCCTTCTTATCCAAATTCCTCACTGCTGCCAAGAAGCATCTTAA	1355
Qy	1164	GACGACGACAAAGCCCGAGCTTCATCCAGATCCGAGGCTGGCCAGGAGCAACCCCTAAG	1223
Db	1356	ACTTGGGTACATTACATTTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTAGTCTC	1415
Qy	1224	ACCTGGGTGCTACTACATCGCCCGCGAGGAGGAGGACTGGGACTACGCCCCGCTAGTCTG	1283
Db	1416	GCCCCGATGACAGAAGTTATAAAAGTCAATATTTGAACAAATGGCCCTCAGCGGATTGGT	1475
Qy	1284	GCCCCGACGACCGCAGCTACAAGAGCCAGTACCTGAACAAACGGCCCCCAGCGGCA	1343
Db	1476	AGGAAGTACAAAAAAGTCCGATTTATGCGATACACAGATGAAACCTTTAAGACTCGTGAA	1535
Qy	1344	CGCAAGTACAGAAGGTGCGCTTCATGGCCTACCCGACGAGACTTTCAGACCCCGGAG	1403
Db	1536	GCTATTCAGCATGAATCAGGAATCTTGGGACCTTTACTTTATGGGGAAGTTGGAGACACA	1595
Qy	1404	GCCATCCAGCAGAGTCCGGCATCTCTCGCCCCCTGCTGTACGGCGAGGTGGGCGACACC	1463
Db	1596	CTGTTGATTATATTTAACAATCAAGCAAGCAGACACATATAACATCTACCCCTCACGGAA	1655
Qy	1464	CTGCTGATCATCTTCAAGAACCCAGGCCACGAGGCCCTACAACATCTACCCCCACGGCAT	1523

Qy 1764 CAGCGGCGCAACAGATCATGAGCACAAGCGCAACGTGATCCTGTTACGGCTGTTCGAC 1823

Db 1907 GAGAACCGAAGCTGGTACCTCACAGAGAAATATACAACGGCTTTCTCCCAATCCAGCTGGA 1966

Qy 1824 GAGAACCGCAGCTGGTATCTGACCGAGAAACATCCAGGGCTTCTGCCCCAACCCGCTGGC 1883

Db 1967 GTGCAGCTTGAGCATCCAGAGTTCCAAGCCTCCAAACATCATGCACAGCATCAATGGCTAT 2026

Qy 1884 GTGCAGCTGGAAGATCCCGAGTTCAGGGCCAGCAACATCATGCACAGCATCAACGGCTAC 1943

Db 2027 GTTTTTCATAGTTTGCAGTTTCTCAGTTTGTGTCATCAGGTGGCATACTGGTACATTCTA 2086

Qy 1944 GTGTTCCACAGCCTGCAGTGCAGCTGAGCGTGTGCCTGCATCAGGTGGCCTACTGCTACCTG 2003

Db 2087 ACCATTGGAGCAGACACTGACTTCCTTTCTCTGCTTCTCTCTGGATATACCTTCAACAC 2146

Qy 2004 AGCATCGCGGCCAGACCGACTTCCTGAGCGTGTCTTCTCCGGGTATACCTTCAAGCAC 2063

Db 2147 AAAATGGTCTATGAAGACACACTCACCCCTATTCCCAATTCTCAGAGAAACTGTCTTCATG 2206

Qy 2064 AAGATGGTGTACGAGGACACCCCTGACCCCTGTTCCTTCTCCGGCGAGACTGTGTTCA TG 2123

Db 2207 TCGATGGAACCCAGGTCATGGATTCTGGGGTGCACAACTCAGACATTTCCGGAACAGA 2266

Qy 2124 TCTATGGAGAACCCCGCCTGTGGATTCTGGGCTGCCACAACAGCGACTTCGCGAACCGC 2183

Db 2267 GGCATGACCGCCTTACTGAAGGTTTCTAGTTGTGACAGAACACTGGTGATTATTACGAG 2326

Qy 2184 GGCATGACTGCCCTGCTGAAGTCTCCAGCTCCGACAAGAACCGCGGACTACTACGAG 2243

Db 2327 GACAGTTATGAAGATATTTTCAGCATACTTGGCTGAGTAAACAACTGCCATTGAACCAAGA 2386

Qy 2244 GACAGCTACGAGGACATCTCCGCCCTACCTGTGTCTCCAAGAACACGCCATCGAGCCCCGC 2303

Db 2387 ACCTTCTCCCAGAAATTCAGACACACCGTAGCACTAGGCAAGAAAGCAATTTAATGCCACC 2443

Qy 2304 TCCTTCTCCAAAACTCCCGCCACCCCGCAGCACGCGTTCAGAAGCAGTTTCAACGCCACC 2360

RESULT 12

LOCUS I71409 8967 bp DNA PAT 23-DEC-1997

DEFINITION Sequence 1 from patent US 5681746.

ACCESSION I71409

NID g3007544

KEYWORDS .

SOURCE Unknown.

ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 8967)

AUTHORS Bodner,M., De Polo,N.J., Chang,S., Hsu,D.Chi-Tang and Respass,J.G.

TITLE Retroviral delivery of full length factor VIII

JOURNAL Patent: US 5681746-A 1 28-Oct-1997;

FEATURES Location/Qualifiers

source 1. .8967

/organism="unknown"

BASE COUNT 2841 a 1898 c 1833 g 2395 t

ORIGIN

Query Match 24.3%; Score 1083; DB 18; Length 8967;

Best Local Similarity 73.8%; Pred. No. 0.00e+00;

Matches 1680; Conservative 0; Mismatches 597; Indels 0; Gaps 0;

Db 167 GCACCCAGAAGATACTACCTGGGTGCAGTGGAACTGTCTATGGGACTATATGCAAACTGAT 226

Qy 84 GCCACCCCGCTACTACTTGGGCGCGGTGGAGTGTCTCTGGACTACATGCAGAGCGAC 143

Db 227 CTCGGTCAGCTGCCTGTGACCGCAAGATTTCCTCCTAGAGTCCCAAAATCTTTTCCATTG 286

Qy 144 CTGGGCGAGCTCCCGTGGACGCCCGCTTCCCCCGCCGCTGCCAAGAGCTTCCCCTTC 203

Db 287 AACACCTCAGTCGTGTACAAAACACTCTGTTTGTAGAAATTCACGGATCACCTTTTCAAC 346

Qy 204 AACACCAGCGTGGTGTACAAGAAAAACCTGTTCGTGGAGTTCAACGACCACCTGTTCAAC 263

Db 347 ATCGCTAAGCCRAAGGCCACCCCTGGATGGGTCTCTGCTAGGTCTTACCATCCAGGCTGAGGTT 406

Qy 264 ATTGCCAAGCCGCGCCCGCTTGGATGGGCTGTGTGGGCCACCATCCAGGCCGAGGTG 323

Db 407 TATGATACAGTGTCTATTACACTTAAGAACATGGCTTCCCATCCTCTCAGTCTTTCATGCT 466

Qy 324 TACGACACCGTGGTGATCACCTTGAAGAACATGGCCAGCCACCCCGTCAGCCTGCACGCC 383

Db 467 GTTGGTCTATCCTACTGGAAGCTTCTGAGGGAGCTGAATATATGATGATCAGACCACTCAA 526

Qy 384 GTGGCGTGAGCTACTGCAAGGCCAGCGAGGGCGCCGAGTACGACCGACGAGCTCCGAG 443

Db 527 AGGAGAGAAAGATGATAAAGTCTTCCCTGGTGGAAAGCCATACATATGTCGTGGCAGGTC 586

Qy 444 CCGAGAAAGGAGCAGCAAGGTGTTCCCGGGGGGAGGCCACACCTACGTGTGGCAGGTG 503

Db 587 CTGAAAGAGAAATGGTCCAAATGGCCTCTGACCCACTGTGCTTACCTACTCATATCTTTCT 646

Qy 504 CTTAAGCAGAACCGCCCTATGGCCAGCGACCCCTGTGCTGACCTACAGCTACCTGAGC 563

Db 647 CATGTGCACCTGCTAAAGACTTGAAATTCAGGCCCTCATTGGAGGCCCTACTACTATGTAGA 706

Qy 564 CACGTGCACCTGCTGAAGGATCTGAACAGCGGGTGTATCGCGCCCTGCTGCTGTGTGCGC 623

Db 707 GAAGGGACTCTGSCCAAGGAAAGTTGGCACTCAGAAACAAAGAACTCCTTGATGCAGGATAGG 766

Qy 624 GAGGGCAGCCTGGCCCAAGGAGAAAAACCCAGACCCCTGTCACAAAGTTTCCTGCTGTTCGCC 683

Db 767 GTATTTGATGAAGGGAAAGTTGGCACTCAGAAACAAAGAACTCCTTGATGCAGGATAGG 826

Qy 684 GTGTTCCAGCAGGGGAAAGCTGGCACAGCGAGACTAAGAACAGCCTGATGCAGGACCGC 743

Db 827 GATGCTGCATCTCTCGGGCCTGGCCTAAAATGCACACAGTCAATGGTTATGTAAACAGG 886

Qy 744 GACGCCGCCAGCGCCCGCCTGGCCCCAAGATGCACACCCGTTAACGGCTACGTGAACCGC 803

Db 887 TCTCTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATG 946

Qy 804 AGCCTGCCCGGCTGATCGGCTGCCACCGCAAGACGCTGTACTGGCAGCTCATCGGCATG 863

Db 947 GGCACCACTCCTGAAGTGCACTCAATATTCTCTGAAGGTCACACATTTCTTGTGAGGAAC 1006

Qy 864 GGCACACCCCTGAGGTGCACAGCATCTTCTCTGGAGGGCCACACCTTCTCTGGTGGCGAAC 923

Db 1007 CATGCGCAGGCTCTTGAAATCTCGCCAATAACTTTCCTTACTGCTCAAACTCTTTG 1066

Qy 924 CACCGCCAGCGCCTGGAGATCAGCCCCATCCTTCTCTGACTGCCCCAGACCCCTGCTG 983

Db 1067 ATGGACCTTGGACAGTTTCTACTGTTTGTGATATCTCTTCCCACCAACATGATGGCATG 1126

Qy 984 ATGGACCTAGGCCAGTTCTCTGCTGTTCTGCCACATCAGACGCCACCCAGCAGCGCATG 1043

Db 1127 GAAGCTTATGTCAAAGTACACAGCTGTCCAGAGGNAACCCCAACTACGAATGAAAAATAAT 1186

Qy 1044 GAGGCTTACGTGAAGGTTGGACAGCTGCCCCGAGGAGCCCGACGTGGCATGAAGAAAC 1103
Db 1187 GAAGACGGGAACACTATGATGATCTTACTGATTTCTGAATGGATGCTGGTCAGGTTT 1246
Qy 1104 GAGGAGCCGAGCACTACGACGACCACTGACCGACAGCGAGATGGATGTCGTAGGCTTC 1163
Db 1247 GATGATCACAACTCTCCTTCCTTTATCCAAATTCGCTCAGTTGCCAAGAAGCATCCTAA 1306
Qy 1164 GACGACCAACAGCCCCAGCTTCATCCAGATCCGAGCGTGGCCCAAGAAGCAACCTAAG 1223
Db 1307 ACTTGGGTACATTACATTTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTACTCCTC 1366
Qy 1224 ACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCGCTAGTACTG 1283
Db 1367 GCCCCCGATGACAGAAGTTATAAAGTCAATATTTGAACAATGGCCCTCAGCGGATTGGT 1426
Qy 1284 GCCCCCGACGACCGCAGCTACAAGGCCAGTACCTGAAACAACGGCCCCCAGCGCATCGGC 1343
Db 1427 AGGAAGTACAAAAGTCCGATTTATGGCATACACAGATGAAGACCTTTAAGACTCGTGAA 1486
Qy 1344 CGCAAGTACAAGAGGTGCGCTTCATGGCCTACACCGACGAGACTTTCAAGACCCGCGAG 1403
Db 1487 GCTATTGAGCATGAATCAGGAATCTTGGGACCTTTACTTTATGGGGAAGTTGGACACACA 1546
Qy 1404 GCCATCCAGCAGACTCCGGCATCCTCGGCCCTGCTGTACGGCGAGGTGGCGACACC 1463
Db 1547 CTGTTGATTATATTAAAGATCAAGCAAGCAGACCATATAACATCTACCTCAGCGAATC 1606
Qy 1464 CTGCTGATCATCTTCAAGAACCCAGCCAGCGCCCTACAAACATCTACCCCCACGGCATC 1523
Db 1607 ACTGATCTCGTCTTGTATTCCAGGAGATTACCAAAAGTGCTAAACATTTGAAGGAT 1666
Qy 1524 ACCGACGTGGCCCCCTGTACAGCGCCGCTGCCCAAGGGCGTGAAGACCTGAAGGAC 1583
Db 1667 TTTCCAATTCTGCCAGGAGAAATATTCAAATATAAATGGACACTGACTGTAGAACATGGG 1726
Qy 1584 TTCCCCATCTCGCCCGCGAGATCTTCAAGTACAAGTGGACCGCTGACCGTGGAGGACGGC 1643
Db 1727 CCAACTAAATCAGATCCTCGGTGCTGACCGCTATTACTCTAGTTTCTGTAATATGGAG 1786
Qy 1644 CCCACCAAGAGCACCCCGCTGCTGACCCGCTACTACAGCAGCTTCGTGAACATGGAG 1703
Db 1787 AGAGATCTAGCTTCAGGACTCATTTGGCCCTCTCCTCATCTGTACAAAGAAATCTGTAGAT 1846
Qy 1704 CGCGACTGGCCTCCGGACTGATCGGCCCTGCTGATCTGTCTACAGGAGCGGTGGAC 1763
Db 1847 CAAAGAGGAACCAAGATAATGTACACAAGAGGAATGCTCATCTGTTTCTGTATTTGAT 1906
Qy 1764 CAGCGCGCAACCAAGATCATGAGCGCAACGTAACCTGTTTTCAGCGTGTTCGAC 1823
Db 1907 GAGAACCGAAGCTGGTACCTCACAGAGAAATATACAACGCTTTCTCCCAATCCAGCTGGA 1966
Qy 1824 GAGAACCGCAGCTGGTATCTGACCGAGAAATCCAGGCTTCTGCCCCAACCCCGCTGGC 1883
Db 1967 GTGCAGCTTGAAGTCCAGAGTTCCAAGCCTCCAAGCATCATGCACAGCATCAATGGCTAT 2026
Qy 1884 GTGCAGCTGGAAGATCCCAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGCTAC 1943
Db 2027 GTTTTTGATAGTTTGCAGTTGTCACTTTGTTGATGAGGTGGCACTACTGGTACATTCTA 2086
Qy 1944 GTGTTCCAGAGCCTGCAGCTGAGCGTGTGCTGCATGAGGTGGCCTACTGGTACATCCTG 2003
Db 2087 AGCATTGGAGCAGACGACTGACTTCCTTTCTCTCTCTGGAATATACCTTCAACAC 2146

Qy 2004 AGCATCGCGCCAGACAGCACTTCCTGAGCGTGTCTTCTCCGGGTATACCTTCAAGCAC 2063
Db 2147 AAAATGCTCTATCAAGACACACTCACCCCTATTCCTCATTTCTCAGGAGAAACTGTCTTCATG 2206
Qy 2064 AAGATGCTGTACGAGGACACCCCTGACCCCTGTTCCTCTTCCGGCGAGACTGTGTTTCATG 2123
Db 2207 TCGATGCAAAACCCAGGTCTATGGATTCTGGGGTCCACAACTCAGACTTTTCGGAACAGA 2266
Qy 2124 TCTATGAGAAACCCCGCTGTGGATTCTGGGCTGCCACAACAGCGACTTCGCAACCGC 2183
Db 2267 GGCATGACCCGCTTACTCAAGGTTTCTAGTTGTGACAAGACACTGCTGATTATTACGAG 2326
Qy 2184 GGCATGACTGCCCTGCTGAAAGTCTCCAGTGGACAAAGAACACCGGCGACTACTACGAG 2243
Db 2327 GACAGTTATGAAGATATTTACGACATACCTTCTGCTGACTAAACAATGCCATTGAACCAAGA 2386
Qy 2244 GACAGCTACGAGACATCTCCGCTACCTGCTGTCCAAGAAACAACGCCATCGAGCCCGC 2303
Db 2387 ACCTTCTCCCAAGATTCGAAGACACCCGTAGACACTAGGCAAAAGCAATTTAATGCCACC 2443
Qy 2304 TCCTTCTCCAAAACTCCCGCCACCCCGACGCGCTCAGAAGCAGTTCAACGCCACC 2360

RESULT 13
LOCUS I31901 9009 bp DNA PAT 20-DEC-1996
DEFINITION Sequence 3 from patent US 5583209.
ACCESSION I31901
NID g1822692
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 9009)
AUTHORS Lollar,J.S. and Runge,M.S.
TITLE Hybrid human/porcine factor VIII
JOURNAL Patent: US 5583209-A 3 10-DEC-1996;
FEATURES Location/Qualifiers
source 1..9009
/organism="unknown"
BASE COUNT 2853 a 1907 c 1844 g 2405 t
ORIGIN

Query Match 24.3%; Score 1083; DB 18; Length 9009;
Best Local Similarity 73.8%; Pred. No. 0.00e+00;
Matches 1680; Conservative 0; Mismatches 597; Indels 0; Gaps 0;

Db 208 GCCACCAGAGATACTACTGCGTGCAGTGGAACTGTCTATGGGACTATATGCAAAAGTCAT 267
Qy 84 GCCACCCCGCTACTACTGCGCGCCGTGGAGTGTCTCTGGGACTACATGCAAGCGCAC 143
Db 268 CTCGGTCAGCTGCCTGTGGACGCAAGATTTCCTCCTAGAGTGCCTAAATCTTTTCCATTTC 327
Qy 144 CTGGCGGAGCTCCCCGTGGACGCCCCGCTTCCCCCCCCCGCTGCCCAAGAGCTTCCCCCTTC 203
Db 328 AACACCTCAGTCGTGTACAAAAGAACTCTGTTTGTAGAAATTCACGGTTTCACCTTTTCAAC 387
Qy 204 AACACCAGCGTGGTGTACAAGAAAACCCCTGTTCTGTGGAGTTCAACCAACCACTGTTCAAC 263
Db 388 ATCGCTAAGCCAAAGCCACCCCTGGATGGGTCTGTAGGTCTTACCATCCAGGCTGAGGTT 447
Qy 264 ATTGCCAAAGCCCGCCCCCTTGGATGGGCTGCTGGGCCCCCAACCATCCAGGCCGAGGTG 323
Db 448 TATGATACAGTGGTCATTACACTTAAGAACATGGGTGCCCATCCTCTCAGTCTTCATGCT 507

Jul 9 16 17

US-08-717-294-42 page 10

Db	1948	GAGAACCGAAGCTGGTACCTCACAGAGAAATATACAGCGTTTCTCCCAATCCAGCTGA	2007
QY	1824	GAGAACCGAGCTGGTATCTGACCGAGAAATCCAGCGCTTCTGCCAACCCCGGTGGC	1883
Db	2008	GTGCAGCTTGAGGATCCAGAGTTCGAAGCCTCCAACATCATGCACAGCATCAATGGCTAT	2067
QY	1884	GTGCAGCTGGAGAATCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGCTAC	1943
Db	2068	GTTTTTGATAGTTTGCAGTTGTCAAGTTTGTTCATGAGGTGGCATACTGGTACATTCTA	2127
QY	1944	GTGTTCCGACAGCTGCAGCTGAGCGGTGTGCCTGCATGAGGTGGCTACTGGTACATCTGT	2003
Db	2128	AGCATTTGGAGCACAGACTGCATTCTCTTCTCTCTTCTCTCGATATACCTTCAAACAC	2187
QY	2004	AGCATCGGGGGCCAGACCGCACTTCCCTGAGCGGTGTTCTTCTCCGGGTATACCTTCAAGCAC	2063
Db	2188	AAATFGTCTATGAAGACACACACTCACCCCTATTTCCCATTTCTCAGGAGAACTGCTTTCATG	2247
QY	2064	AGATGGTGTAGGAGGACACCCCTGACCCCTGTTCCCTTCTCCGGCGAGACTGTGTTTCATG	2123
Db	2248	TCGATGGAAACCCAGGCTCTATGGATTCTGGGTTGCCACAACATCAGACTTTCGGAACAGA	2307
QY	2124	TCTATGGAGAACCCCGGCTGTGGATTCTGGGCTGCCACAACAGCGACTTCCCGCAACCGC	2183
Db	2308	GGCATGACCGCCTTACTCAAGGTTTCTAGTTGTGACAAGAACTGGTGATTATTACGAG	2367
QY	2184	GGCATGACTGGCCTGCTGAAAGTCTCCAGTCTCCAGTCCGCAAGAACACCCGGCGACTACTAGGAG	2243
Db	2368	GACAGTTATGAAGATATTTCAGCATACCTTGCTGAGTAAAAACAATGCCATTGAACCAAGA	2427
QY	2244	GACAGCTACGAGGACATCTCCGGCTACCTGCTGTCCAGAACAGCGCCACTCGAGGCCCGC	2303
Db	2428	AGCTTCTCCAGAAATTCAGACACCCCTAGCACTAGGCAAAAGCAATTTAATGCCACC	2484
QY	2304	TCCTTCTCCCAAAACTCCGGCCACCCAGACGCGTCAAGAGCAAGCAATTCACGCCCAAC	2360

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 377.

FEATURES

source
1. .378
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="785742"
/clone_lib="Soares total fetus Nb2HF8 9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
<1. .>378
/db_xref="GDB:5983361"
BASE COUNT 90 a 108 c 100 g 80 t
ORIGIN

Query Match 1.3%; Score 56; DB 7; Length 378;
Best Local Similarity 62.8%; Pred. No. 2.22e-41;
Matches 137; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Db 41 GGAGTTTGTGGTAACTGGAACAAAAACGGGTGCATGTCAACCTGTTTGAGACCCCTGT 100
|| ||| | || ||| | ||| | || | | ||| | || | ||| |
QY 3815 GGTGTTCTTCGGCAACGTGGACAGCGGCATCAAGCACACATCTTCAACCCCCCGAT 3874

Db 101 GGAGGCTCAGTACGTGAGATTGTACCCACAGCTGCCACAGGCGCTGCACCTCTGCGCTT 160
|| | ||| | | || ||||| || | || | ||| ||||| |
QY 3875 CATCGCCCGCTACATCCGCCTGGACCCACCCACATACAGCATCCGGACACCCCTGCGCAT 3934

Db 161 TGAGCTACTGGCTGTGAGCTGAACGGATGCGCCCAATCCCTGGCGCTGAAGAATAACAG 220
||||| ||||| || ||||| | || | || ||||| || ||| ||
QY 3935 GGAGCTGATGGCTCGACCTGAACAGCTGCAGCATGCCCTGGGCATGGAGCAAGGC 3994

Db 221 CATCCCTGACAAGCAGATCAGCGGCTCCAGCAGCTACA 258
|||| ||| ||||| ||||| ||||| || ||
QY 3995 CATCAGCAGCGCCAGATCACCGCCTCCAGCTACTTCA 4032

RESULT 7
LOCUS N39692 522 bp mRNA EST 22-JAN-1996
DEFINITION yx92a03.r1 Homo sapiens cDNA clone 269164 5' similar to PIR:A47285 A47285 milk fat globule protein - human ;.
ACCESSION N39692
NID g1163237
KEYWORDS EST.
SOURCE human clone=269164 primer=T7 library=Soares melanocyte 2NbHM vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGCGCAGTTTTTTTTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library

constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.

ORGANISM

Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 522)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE
JOURNAL
COMMENT

The WashU-Merck EST Project
Unpublished (1995)

CONTACT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 426
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source
1. .522
/organism="Homo sapiens"
/clone="269164"
<1. .>522
BASE COUNT 125 a 139 c 146 g 112 t
ORIGIN

Query Match 1.3%; Score 56; DB 19; Length 522;
Best Local Similarity 62.8%; Pred. No. 2.22e-41;
Matches 137; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Db 67 GGAGTTTGTGGTAACTGGAACAAAAACGGGTGCATGTCAACCTGTTTGAGACCCCTGT 126
|| ||| | || ||| | ||| | || | | ||| | || | ||| |
QY 3815 GGTGTTCTTCGGCAACGTGGACAGCGGCATCAAGCACACATCTTCAACCCCCCGAT 3874

Db 127 GGAGGCTCAGTACGTGAGATTGTACCCACAGCTGCCACAGGCGCTGCACCTCTGCGTT 186
|| | ||| | | || ||||| || | || ||||| || | ||| ||||| |
QY 3875 CATCGCCCGCTACATCCGCCTGCACCCACCCACTACAGCATCCGAGCACCCTGCCAT 3934

Db 187 TGAGCTACTGGCTGTGAGCTGAACGGATCGGCAATCCCTGGGCTGAAGAATAACAG 246
||||| ||||| || ||||| | ||| || ||||| || ||| ||
QY 3935 GGAGCTGATGGCTGGCACCTGAACAGCTGCAGCATGCCCTGGGCATGGAGACCAAGCC 3994

Db 247 CATCCCTGACAAGCAGATCACGGCCTCCAGCAGCTACA 284
|||| ||| ||||| ||||| ||||| || ||
QY 3995 CATCAGCAGCGCCAGATCACCGCCTCCAGCTACTTCA 4032

RESULT 8
LOCUS AA165976 555 bp mRNA EST 11-FEB-1997
DEFINITION mr92d06.r1 Stratagene mouse embryonic carcinoma (#937317) Mus musculus cDNA clone 604907 5' similar to gb:M38337 Mouse milk fat globule membrane protein E8 mRNA, complete (MOUSE);, mRNA sequence.
ACCESSION AA165976
NID g1744506

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 555)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:370339
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 335.
Location/Qualifiers
1. .555
/organism="Mus musculus"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. p19 cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'
/db_xref="taxon:10090"
/clone="604907"
/clone_lib="Stratagene mouse embryonic carcinoma (#937317)"
/dev_stage="P19 cell line (ATCC CRL1825)"
/lab_host="SOLR (kanamycin resistant)"
<1. .>555

mRNA
BASE COUNT 140 a 139 c 161 g 113 t 2 others
ORIGIN

Query Match 1.3%; Score 60; DB 23; Length 555;
Best Local Similarity 63.8%; Pred. No. 4.19e-47;
Matches 139; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Db 73 GGAGTTTTTGGGTAACCTGGACAACAAGCCTGAAGGTTAAACATGTTCAACCCGACTCT 132
|| ||| || || ||| ||||| || || || ||| ||||| ||||| || ||
Qy 3815 GGTGTCTTCGGCAACGTGGACAGCAGCGGCATCAAGCACACATCTTCAACCCCCCAT 3874

Db 133 GGAGGCACAGTACATAGGCTGTACCCCTGTTTCGTGCCACCGCGGTGCACCCCTCCGCTT 192
|| | ||||| | ||| |||| | | | | | || ||||| ||| |
Qy 3875 CATCGCCCGTACATCCGCCCTGCACCCCAACCCACTACAGCATCGGCAGCACCCCTGCGCAT 3934

Db 193 CGAGTCCCTGGGCTGTGAGTTGCACGGATGTTCTGAGCCCTGGCCTGAAGATAACAC 252
||||| ||||| || || || || || ||||| || || || || || || ||
Qy 3935 GGAGCTGATGGCTGCCACCTGAACAGCTGCAGCATGCCCTGGGCATGGAGCAAGGC 3994

Db 253 AATTCTGCAGAGCCAGATGTCAGCCTCCAGCAGCTACA 290
|| ||| ||||| | ||||| || || ||
Qy 3995 CATCAGCAGCGCCAGATCACCGCCTCCAGCTACTCA 4032

RESULT 9
LOCUS AA512671 613 bp mRNA EST 08-JUL-1997
DEFINITION vg39g02.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone 863762 5' similar to gb:M38337 Mouse milk fat globule membrane protein E8 mRNA, complete (MOUSE);, mRNA sequence.

ACCESSION AA512671
NID g2250525
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 613)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:507850
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 273.
Location/Qualifiers
1. .613
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer {5' TGTTACCAATCTGAAGTGGAGCGCGCGGAATGGTTTTTTTTTTTTTTTTTTT T 3'}; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="863762"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
<1. .>613

mRNA
BASE COUNT 151 a 163 c 174 g 125 t
ORIGIN

Query Match 1.3%; Score 58; DB 7; Length 613;
Best Local Similarity 63.3%; Pred. No. 3.14e-44;

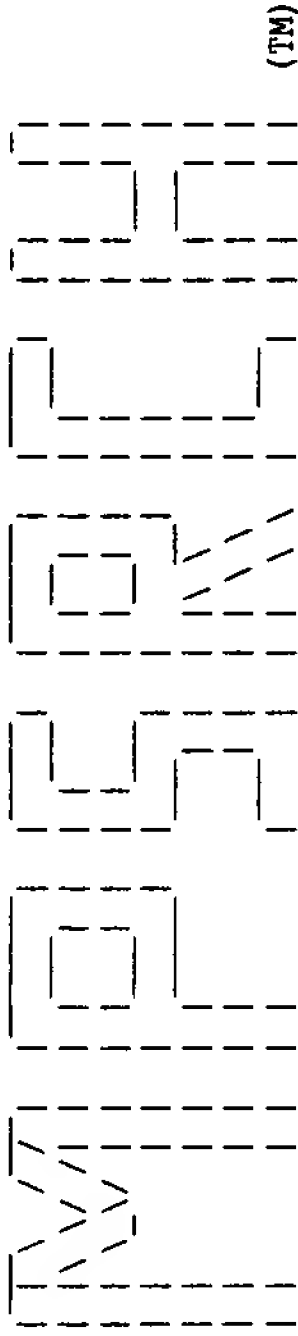
Genexpress_sequence_idt: ylc-0ab01.
FEATURES
source
1. .261
/organism="Homo sapiens"
/isolate="muscular atrophy patient"
/db_xref="taxon:9606"
/clone_lib="normalized infant brain cDNA from B.Soares,
Psychiatry Dept. Columbia University USA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
BASE COUNT 57 a ~79 c 69 g 52 t 4 others
ORIGIN
Query Match 1.2%; Score 53; DB 21; Length 261;
Best Local Similarity 64.2%; Pred. No. 3.71e-37;
Matches 115; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Db 14 CAACCTGTTTGACACCCCTGTGGAGGCTCAGTACGTGAGATTNTATCCCGACGAGCTGCCA 73
||||| ||| | ||||| | ||| ||||| | | ||||| |||
QY 3854 CAACATCTTCAACCCGCCCATCATGCCCGCTACATCGCGCTGACCCCGCTGACCCCGCTACAG 3913
Db 74 CACGGCTGCATCTCGCGTTTGAGCTACTGGGNTGTGAGCTGAACGGATGCGGCAATCC 133
|| | ||||| ||||| | ||||| ||||| || ||||| ||||| || ||
QY 3914 CATCCGGAGCACCTGCGCATGGAGCTGATGGGCTGGACCTGAACAGCTGCAGCATGCC 3973
Db 134 CCTGGGCTGAACAATACAGCATCCCTGACAAGCAGATCAGCGCCTCCAGCAGCTACA 192
||||||| || |||| || |||| |||| ||||| ||||| ||||| || ||
QY 3974 CCTGGGCATGGAGAGCAAGCCCATCAGCGGACGCCCGCAGATCAGCGCCTCGAGCTACTTCA 4032
RESULT 12
LOCUS HSC3AG071 317 bp RNA EST 21-SEP-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-3ag07, mRNA sequence.
ACCESSION F12497
NID g708490
KEYWORDS EST; partial cDNA sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 317)
AUTHORS Genexpress.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) Genethon, B.P. 60, 91002 Evry Cedex France
and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
B.P. 8, 94801 Villejuif Cedex France. E-mail:
genexpress@genethon.fr
2 (bases 1 to 317)
Genexpress.
TITLE The Genexpress cDNA program
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 317)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Cloning method: total mRNA was oligo-(dfr) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
vector;

Sequencing_method: single read, full automatic;
Primer: M13_reverse
cDNA sequence colinear to mRNA
Stretch_removed: nothing
Normalization_method: Bento Soares, P.N.A.S. 91:9228-9232(1994);
Genexpress_library_idt: C;
Genexpress_sequence_idt: ylc-3ag07.
FEATURES
source
1. .317
/organism="Homo sapiens"
/isolate="muscular atrophy patient"
/db_xref="taxon:9606"
/clone_lib="normalized infant brain cDNA from B.Soares,
Psychiatry Dept. Columbia University USA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
BASE COUNT 83 a ~84 c 83 g 67 t
ORIGIN
Query Match 1.2%; Score 55; DB 20; Length 317;
Best Local Similarity 63.0%; Pred. No. 5.77e-40;
Matches 133; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Db 102 GGAGTTGTGGTAACTGGAACAAAAACGGGTGCATGTCAACCTGTTTGGACACCCCTGT 161
||||| | ||||| | ||||| || | | ||||| ||| | |||||
QY 3815 GGTGTTCTTCGGCAACGTGGACAGCAGCGGCATCAAGCAACAATCTTCAACCCCCCAT 3874
Db 162 GGAGGCTCAGTACGTGAGATTGTACCCCGACGAGCTGCCACGGCTGCACTCTGCGCTT 221
|| | |||| | | ||||| || | || || || ||||| ||||| |
QY 3875 CATCGCCCGCTACATCCGCCTGCACCCCGCCACTACAGCATCCGAGCACCCCTGCCGAT 3934
Db 222 TGAGCTACTGGGCTGTGAGCTGAACGGATGCGGCAATCCCTGGGCTGAAGAATAACAG 281
||||| ||||| || ||||| || || | ||||| || || || ||
QY 3935 GGAGCTGATGGGCTGCGACCTGAACAGCTGCAGCATGCCCTGGGCATGGAGACGAAGGC 3994
human ;, mRNA sequence.
Db 282 CATCCCTGACAAGCAGATCAGCGCCTCCAGC 312
|||| | || ||||| ||||| ||||| |||||
QY 3995 CATCAGGACGCCCGCAGATCAGCGCCTCCAGC 4025
RESULT 13
LOCUS W86071 328 bp mRNA EST 17-OCT-1996
DEFINITION zd91b11.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
356829 5' similar to PIR:A47285 A47285 milk fat globule protein -
human ;, mRNA sequence.
ACCESSION W86071
NID g1398676
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project

Jul9 19:46

US-08-717-294-42.mrg

1



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Jul 9 19:29:09 1998; MasPar time 581.39 Seconds
Tabular output not generated. 975.051 Million cell updates/sec

Title: >US-08-717-294-42
Description: (1-4451) from US08717294.seq
Perfect Score: 4451
N.A. Sequence: 1 AAGCTTAAACCATGCCCATG.....ACCTGTACTGAGCGGCGGC 4451
Comp: TTCGAATTGGTACGGGTAC.....TGGACATGACTTCGCCCGCGC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0
Searched: 176886 seqs, 63680241 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq31-2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37

Statistics: Mean 10.111; Variance 7.080; scale 1.428

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2165	48.6	9354	37	T73164 cDNA encoding human B	0.00e+00
2	2111	47.4	4629	13	Q76016 B-domain deleted Fact	0.00e+00
3	1643	36.9	4275	1	N90654 DNA encoding 740 Arg-	0.00e+00
4	1642	36.9	4275	1	N80446 Modified factor VIII:	0.00e+00
5	1640	36.8	4272	1	N80447 Modified factor VIII:	0.00e+00

Jul9 19:46

US-08-717-294-42.mrg

2

6	1090	24.5	4545	1	N80444	Modified factor VIII:	0.00e+00
7	1085	24.4	6300	17	T03571	Factor-VIII cDNA.	0.00e+00
8	1085	24.4	7053	33	T51357	Factor VIII:C coding	0.00e+00
9	1085	24.4	8241	1	N81096	cDNA sequence encodin	0.00e+00
10	1085	24.4	8241	1	N81439	Factor VIII cDNA inse	0.00e+00
11	1085	24.4	8975	3	N50054	Human factor VIII gen	0.00e+00
12	1083	24.3	5035	30	T69811	Factor VIII-dB695-HCI	0.00e+00
13	1083	24.3	7056	10	Q66615	Sequence of human fac	0.00e+00
14	1083	24.3	8967	21	T31031	Factor-VIII full-leng	0.00e+00
15	1083	24.3	9009	33	T61548	Factor VIII:C (Arg168	0.00e+00
16	1083	24.3	9009	8	Q50185	Human Factor VIII cDN	0.00e+00
17	1075	24.2	4616	1	N81545	Human Factor VIII-C a	0.00e+00
18	1079	24.2	7059	1	N81543	Human Factor VIII-C a	0.00e+00
19	1075	24.2	7440	2	N60689	Sequence encoding hum	0.00e+00
20	1073	24.1	4830	1	N81544	Human Factor VIII-C a	0.00e+00
21	1027	23.1	7440	3	N50375	DNA sequence encoding	0.00e+00
22	664	14.9	1623	10	Q66616	Sequence of porcine f	0.00e+00
23	497	11.2	3852	3	N50094	DNA encoding Factor-V	2.16e-275
24	487	10.9	1130	8	Q50187	Porcine Factor VIII A	2.61e-269
25	310	7.0	955	1	Q05844	Factor VIII insert of	3.66e-162
26	310	7.0	955	3	N50053	Sequence of pESVDA.S1	3.66e-162
27	112	2.5	669	3	N50105	DNA encoding truncate	2.81e-45
28	112	2.5	15155	3	N50107	DNA encoding Factor-V	2.81e-45
29	78	1.8	1632	32	T73951	HIV-1 gpl20 synthetic	2.76e-26
30	78	1.8	1632	22	T13556	Syngpl20mn.	2.76e-26
31	78	1.8	5337	31	T69892	DNA for aggrecanase a	2.76e-26
32	78	1.8	5337	31	T69893	DNA for aggrecanase a	2.76e-26
33	76	1.7	1459	14	Q93128	XhoI-EagI DNA fragmen	3.40e-25
34	69	1.6	1260	8	Q50186	Porcine Factor VIII p	2.07e-21
35	67	1.5	366	3	N50373	HaeIII insert 34-HI b	2.43e-20
36	67	1.5	690	3	N50372	Fragment contg. porci	2.43e-20
37	68	1.5	1129	25	T45262	Heregulin-alpha fusio	7.10e-21
38	56	1.3	1384	14	Q91199	HMFG 46 kDa antigen p	1.50e-14
39	56	1.3	1934	14	Q91198	HMFG 46 kDa antigen c	1.50e-14
40	59	1.3	2297	24	T33956	Human Factor V heavy	4.12e-16
41	59	1.3	2297	24	T33949	Human Factor V cDNA P	4.12e-16
42	59	1.3	6909	15	T03920	Human Factor-V wild-t	4.12e-16
43	59	1.3	6925	24	T33942	Human Factor V cDNA.	4.12e-16
44	57	1.3	6925	24	T33948	Human Factor V mutant	4.54e-15
45	52	1.2	2296	24	T33950	Human Factor V mutant	1.70e-12

ALIGNMENTS

RESULT	1
ID	T73164 standard; cDNA; 9354 BP.
AC	T73164;
DT	08-APR-1998 (first entry)
DE	cDNA encoding human B-domain deleted factor VIII.
KW	Post-translational regulatory element; PRE; enhancer II; intronless gene;
KW	surface antigen gene; cytoplasmic accumulation; targeted delivery;
KW	near consensus splice sequence; blood coagulation factor; factor VIII;
KW	factor IX; ss.
OS	Homo sapiens.
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	2965..7380
FT	/*tag= a
FT	misc_feature 5165..5174
FT	/*tag= b
FT	/note= "5' near consensus site"
FT	misc_feature 5695..5703
FT	/*tag= c
FT	/note= "5' near consensus site"
FT	misc_feature 6320..6328

FT /tag= d
FT /note= "5' near consensus site"
FT misc_feature 6595..5603
FT /tag= e
FT /note= "5' near consensus site"
FT misc_feature 7045..7053
FT /tag= f
FT /note= "5' near consensus site"
FT misc_feature 7143..7152
FT /tag= g
FT /note= "5' near consensus site"
FT misc_feature 3296..3312
FT /tag= h
FT /note= "3' near consensus site"
FT misc_feature 4798..4817
FT /tag= i
FT /note= "3' near consensus site"
FT misc_feature 5023..5045
FT /tag= j
FT /note= "3' near consensus site"
FT misc_feature 5333..5355
FT /tag= k
FT /note= "3' near consensus site"
FT misc_feature 5520..5538
FT /tag= l
FT /note= "3' near consensus site"
FT misc_feature 5604..5632
FT /tag= m
FT /note= "3' near consensus site"
FT misc_feature 5717..5745
FT /tag= n
FT /note= "3' near consensus site"
FT misc_feature 6239..6258
FT /tag= o
FT /note= "3' near consensus site"
FT misc_feature 6658..6682
FT /tag= p
FT /note= "3' near consensus site"
FT misc_feature 7159..7176
FT /tag= q
FT /note= "3' near consensus site"
FT misc_feature 7196..7209
FT /tag= r
FT /note= "3' near consensus site"
FT misc_feature 7289..7315
FT /tag= s
FT /note= "3' near consensus site"
FT misc_feature 7411..7429
FT /tag= t
FT /note= "3' near consensus site"
FT misc_feature 7611..8197
FT /tag= u
FT /note= "PRE sequence"
PN W09733994-A1.
PD 18-SEP-1997.
PF 10-MAR-1997; U03561.
PR 11-MAR-1996; U5-683839.
PA (IMMU-) IMMUNE RESPONSE CORP.
PI Bidlingmaier S, Ill CR;
DR WPI; 97-470874/43.
DR P-PSDB; W23414.
PT Vector for increased expression of intronless genes - comprises
PT intronless gene with at least one near consensus splice sequence, a
PT promoter and at least one viral cis-acting post-transcriptional

PT regulatory element
PS Example 1; Pages 21-31; 59pp; English.
CC The present sequence represents human B-domain deleted factor VIII
CC cDNA, and a post-translational regulatory element (PRE) of the
CC Hepatitis B virus, which is present 3' of the STOP codon for factor VIII.
CC PRE sequences have been shown to function in cis to increase the
CC steady-state levels of surface gene transcripts by facilitating
CC cytoplasmic accumulation of these transcripts. The present sequence
CC is part of a novel vector, comprising an intronless gene containing
CC 1 or more near consensus splice sequences operably linked to a
CC promoter sequence so that the gene is transcribed in a cell.
CC Intronless gene transcripts which contain near consensus splice site
CC sequences are believed to get tied up in the nucleus of the cell where
CC splicing occurs, rather than being transported to the cytoplasm where
CC they can be translated into proteins. The PRE sequences are transcribed
CC along with the gene, causing export of the gene transcript from the
CC nucleus into the cytoplasm of the cell. The vector can be used
CC to increase the expression of an intronless gene containing at least one
CC near consensus splice sites, preferably cDNA encoding a blood coagulation
CC factor, particularly Factor VIII or IX. The complex allows the targeted
CC delivery of the vector to a specific cell, e.g. hepatocytes when the
CC ligand is an asialoglycoprotein which binds the asialoglycoprotein
CC receptor present on their surface.
SQ Sequence 9354 BP; 2506 A; 2239 C; 2161 G; 2448 T;

Query Match 48.6%; Score 2165; DB 37; Length 9354;
Best Local Similarity 74.8%; Pred. No. 0.00e+00;
Matches 3262; Conservative 0; Mismatches 1097; Indels 0; Gaps 0;

Db 3022 gccaccagaagatactacctgggtgcagtggaaactgtcatgggactatatgcaaaagtgat 3081
Yy ||||| | | ||||| ||||| || ||||| ||||| ||||| ||||| ||||| || ||
84 GCACCCGCCGTACTACTCTGGGCGCGTGGAGCTGTCTCTGGGACTACATGCAGAGCGGAC 143

Db 3082 ctcggtgagctgcctgtggacgaagatttcctctctagatgccaatacttttccattc 3141
Yy || || ||||| || ||||| || || || || || || || || || || || || || ||
144 CTGGGCGAGCTCCCCGTGACGCCCGCTTCCCCCGCGCTGCCAAGAGCTTCCCCCTTC 203

Db 3142 aacacctcagtcgtgtacaaaaagactcgtgtttagaattcacggttcaccttttcaac 3201
Yy ||||| || ||||| || || ||||| || || ||||| || || ||||| ||||| |||||
204 AACACCACCGTGGTGTACAGAAACCCGTGTCTGGAGTTCAACGACCACCGTGTCAAC 263

Db 3202 atcgctaagccaaggccaccctggatgggtctgctaggtcctaceatccaggctgaggtt 3261
Yy || || ||||| || || ||||| ||||| ||||| || || ||||| ||||| |||||
264 ATTGCCAAGCCGCGCCCCCTGGATGGGCTGTCTGGGCGCCACCATCCAGGCCGAGGTG 323

Db 3262 tatgatacagtggtcattacacttaagaacatggttcccatcctctcagttcattcttctgct 3321
Yy || || || ||||| || || || ||||| ||||| || || ||||| || || || || ||
324 TAGGACACCGTGGTGTACCCCTGAAGAACATGGCCAGGCCACCCCTCAGCCTGCACGCC 383

Db 3322 gttggtgtatcctactggaagcttctgagggaggtgaatatgatgatcagaccagtcaa 3381
Yy || || || ||||| || || ||||| || || || || || || || || || || ||
384 GTGGCGCTGAGCTACTGGAAGGCCAGCGGGCGCGGAGTACGACGACGACGCTCCAG 443

Db 3382 agggagaaagaatgataaagtcttccctggtggaagccaatacatatgtctggcaggtc 3441
Yy || ||||| || || || || || || || || || || || || || || || || || || || ||
444 CCGGAGAGGAGGACGACCAAGGTCTTCCCGGGGGGAGCCACACCTACGTTGGCAGGTG 503

Db 3442 ctgaaagagaatggtccaatggcctctgaccactgtgccttacctactcatatctttct 3501
Yy || || ||||| || || ||||| ||||| ||||| ||||| ||||| || || || || ||
504 CTTAAGGAGAACCGCCCCATATGGCCAGCGACCCCTGTGCTGACCTACAGTTACCTGAGC 563

Db 3502 catgtggacctggtaaaagacttgaattcaggcctcattggagccctactagtagtaga 3561
Yy || ||||| ||||| || || || || || || || || || || || || || || || || ||
564 CACGTGGACCTGGTGAAGGATCTGAACACGGGGGTGATCGCGCGCCCTGCTGGTGTGCGC 623

Db 2214 -----acc--a--a---g---agaa--a-t--a-----actc-- 2230
Qy 2304 TCCTTCTCCAAAACTCCGGCCACCCAGCAGCAGCGTCAAGAGGAGTTCAAGCCGACCCCC 2363

Db 2231 ---gtact--a-----ctctt---c-ag-tca---g-----a---t-caa---ga-----gga- 2258
Qy 2364 CCCTGTCTGAAGCGCCACCAAGCGGAGATCACCCGGCACCCCTGCAAGAGGAGGACTTCGACATCTAC 2423

Db 2259 -a-attgactatgatgataccatatcagttggaatgaagaaggaagattttgacatttat 2316
Qy 2424 GAGATCGACTACGACGACACCATCAGCGTGGAGATGAAGAGGAGGACTTCGACATCTAC 2483

Db 2317 gatgaggatgaaaaatcagagcccccgagctttcaaaagaataaacacgacactattttat 2376
Qy 2484 GACGAGGACGAGAACGACAGCCCCCGCTCCTTCCAAAGAAACCCCGCCTACTACTTCATC 2543

Db 2377 gctgcagtgagagggctctgggattatgggatgagtagtctccacatgttctaagaaac 2436
Qy 2544 GCCGCGGTGAGCGCGCTGTGGGACTACGGGATGAGGAGGAGCGCCGCCACGTCCTGCGCAAC 2603

Db 2437 agggctcagagtggcagtgccctcagttcaagaaaagtgttttccagggaatttaactgat 2496
Qy 2604 CGCGCCCCAGCGGAGCGGTGCCCCAGTTCAAGAGGTGGTGTTCAGGAGTTTCACCGAC 2663

Db 2497 ggctcccttactcagcccttataccgtggagaactaaatgaacatttgggactcctgggg 2556
Qy 2664 GGCAGCTTACCCAGCCCCCTGTACCGCGGCGAGCTGAACGAGGACCTGGCGCTGCTCGGC 2723

Db 2557 ccatatataagagcagaagttgaagataaatatcatggtaaccttcagaaaatcaggccctc 2616
Qy 2724 CCCTACATCCGCGCGGAGGTGAGGACAAACATCATGGTGACCTTCCGGCAACCAAGCCTCC 2783

Db 2617 cgtccctattcctctattctagccttattcttatgagaagatcagagcgaagagca 2676
Qy 2784 CGGCGCTACTCCTTCTACTCCTCCCTGATCAGCTACGAGGAGGACAGCGCGGCGGCC 2843

Db 2677 gaacctagaaaaaactttgtcaagcctaataatgaaccaaacttacttttgaaaagtgc aa 2736
Qy 2844 GAGCGCCGCAAGAACTTCGTGAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTGCAG 2903

Db 2737 cateatatggcaccactaaaagatgagtttgactgcaagcctgggcttatttctctgat 2796
Qy 2904 CACCACATGGCCCCCACCNAAGCAGGAGTTTCGACTGCAAGGCCTGGGCCCTACTTTCAGCGAC 2963

Db 2797 gttgacctgaaaaaagatgtgcactcaggcctgattggaccccttctggtctgceacact 2856
Qy 2964 GTGGACCTGGAGAAAGGACGTGCACAGCGGCCCTCATCGGCCCCCTGCTGGTGTGCCACAC 3023

Db 2857 aacacactgaacccctgctcatgaggagacaagtgacagtacagggaatttgctctgtttctc 2916
Qy 3024 AACACCCCTGAACCCCCCCACGAGCTGAAGCTGGTACTTCACCGAGAACTTGGCCCTGTCTTC 3083

Db 2917 accatctttgatgagaccaaaagctggtacttcactgaaaaatatggaaagaaactgcagg 2976
Qy 3084 ACCATCTTCGACGAGACTAAGAGCTGGTACTTCACCGAGAACTTCGAGCGGCAACTGCCGC 3143

Db 2977 gctccctgcaatatccagatggaagatcccacttttaagagagaattatcgcttccatgca 3036
Qy 3144 GCCCCCTGCAACATCCAGATGGAAGATCCCCACCTTCAGGAGAACTTACCGCTTCCAGGCC 3203

Db 3037 atcaatggctacataatggatacactacctggcttagtaatggctcaggatcaaaaggatt 3096
Qy 3204 ATCAACGGGTACATCATGTGACACCCCTGCCCGGCGTGGTGATGGCCCGAGGACCGCATC 3263

Db 3097 cqatggtatctgctcagcatgggcagcaaatgaaaacatccattctattcatttcagtgga 3156
Qy 3264 CGGTGGTACCTGCTGTCTATGGGCAGCAACGAGAACATCCACAGCATCCACTTCAGCGGC 3323

Db 3157 catgtgttcactgtacgaaaaaaagagaggtataaaaatggcactgtacaatctctatcca 3216
Qy 3324 CACGTTTTCACCGTCGCGCAAGAGAGGAGGTACAGATGGCCCTGTACAAACCTGTACCC 3383

Db 3217 ggtgtttttgagacagtggaaaatgttaccatccaaagctggaatttggcgggtggaatgc 3276
Qy 3384 GCGGTGTTCCGAGACTGTGGAGATGCTGCCAGCAGAGCGCCGGGATCTGGCGCGTGGAGTGC 3443

Db 3277 cttattggcgagcatctacatgctgggatgagcacacttttctggtgtacagcaataag 3336
Qy 3444 CTGATCGCGGAGCACCTGCACGCCGGCATGAGCACCCCTGTTCCTGGTGTACAGCAACAAG 3503

Db 3337 tgtcagactcccctgggaatggcttctggacacattagagatttttcagattacagcttca 3396
Qy 3504 TGCCAGACCCCTCTGGGATGGCCAGCGGCCACATCCGGCACTTCCAGATCACCGCCAGC 3563

Db 3397 ggacaatatggacagtgggcccccaagctggccagacttcattattccggatcaatcaat 3456
Qy 3564 GGCAGTACGGCCAGTGGGCTCCCAAGCTGGCCCCGCTGCCTACAGCGCGGAGCATCAAC 3623

Db 3457 gctggagcacaaggagcccttttcttgatcaagtggtgatctgttggcacaatgatt 3516
Qy 3624 GCTGGTCGACCAAGGAGCCCTTCTCTGGATCAAGGTGGACCTGCTGGCCCCCCTATGATC 3683

Db 3517 attcacggcatcaagaccaggtggcccgctcagaagttctccagcctctacatctctcag 3576
Qy 3684 ATCCACGGCATCAAGACCCAGGGCGCCCGCAGAAAGTTCAAGCGCCTGTACATCAGCCAG 3743

Db 3577 ttatcatcatgtatagtcttgatgggaagtggcagacttatcgagggaattccact 3636
Qy 3744 TTCAATCATATGTACTCTCTAGCGGCAAGAAAGTGCAGACCTACCGCGGCAACAGCAC 3803

Db 3637 ggaaccttaatgggtcttcttggcaatgtggattcatctggataaaacacaatatattt 3696
Qy 3804 GGCACCCCTGATGGTGTCTTCGGCAACCTGACAGCAGCGGGCATCAAGCACAAACATCTTC 3863

Db 3697 aaccctccaattattgctcgatacatccgtttgcacccaactcattatagcattcgacgc 3756
Qy 3864 AACCCCCCATCATCGCCCGCTACATCCGGCTGCACCCCAACCCACTACAGCATCCGCAGC 3923

Db 3757 actcttcgcagtgagttgatgggctgtgattaaatagttgcagcatgccattgggaatg 3816
Qy 3924 ACCCTGGCATGGAGCTGATGGGCTGCGACCTGAACAGCTGCAGCATGCCCTTGGGCATG 3983

Db 3817 gagagtaaaagcaatatcagatgcacagattactgttcatctacttaccaatatgttt 3876
Qy 3984 GAGAGCAAGGCCATCAGCGACGGCCAGATCACCGCCTCCAGCTACTTCACCAACATGTTTC 4043

Db 3877 gccacctggtctctcttcaaaagctcgacttcacctcacaaggaggagtaaatgcctggaga 3936
Qy 4044 GCCACCTGGAGCCCCCAGCAAGGCCCGCCTGCACCTCAGGGCGCGCAGCAACGGCTGGCGC 4103

Db 3937 cctcaggtgaataatccaaaagagtggctgcaagtggacttcacagaggaataatgaaatgc 3996
Qy 4104 CCCCAGGTGAACACCCCAAGGAGTGGTGCAGGTGCACCTTCCAGAAAACCATGAAGGTG 4163

Db 3997 acaggagttaactactcagggagtaaaaatctctgttaccagcatgtatgtgaaggagttc 4056
Qy 4164 ACTGGCGTGCACCCCGCGCGTCAAGAGAGCCTGCTGTACCAGCATGTACGTGAAGGAGTTC 4223

Qy 144 CTGGGCGAGCTCCCGGTGGACGCCCGCTTCCCCCGCGGTGCCCCAAGAGCTTCCCTTC 203

Db 184 aacacctcagtcgtgtacaaaaagactctgtttgtagaattcaccggaatcaccttttcaac 243
||||| || ||||| || || ||||| || || ||||| || ||||| || |||||

Qy 204 AACACCAAGCTGTGTACAAAGAACCCCTGTTCGTGGAGTTTCCCGACACCTGTTCAC 263

Db 244 atcgctaagccaagccaccctggatgggtgtgctgagtgctaccatccaggctgaggtt 303
|| || ||||| || || ||||| || || ||||| || || ||||| || || |||||

Qy 264 ATTGCCAAGCGCGCCCCCTGGATGGCTGTCTGGCCCCCACCATTACAGGCCGAGGTG 323

Db 304 tatgatacagtggtcattacacttaagaacatggcttcccatcctgtcagttctcatgct 363
|| || || ||||| || || || ||||| || || || ||||| || || || |||||

Qy 324 TACGACACCGTGTGATCACCCCTGAAGAACATGGCCAGCCACCCTGAGCCTGCACGCC 383

Db 364 gttggtgtatcctactggaagcttctgaggagctgaaatgatgatcagaccagtcac 423
|| || || ||||| || || ||||| || || || || || || || || || || || || ||

Qy 384 GTGGGCTGAGCTACTGGAAGGCCAGCGAGGGCGCCGAGTACGACGACGAGCTCCAG 443

Db 424 agggagaagaagatgataaagtcttccctggtggaagccatacatatgtctgaggtc 483
||||| || || || || ||||| || || ||||| || || ||||| || || || |||||

Qy 444 CGGAGAAGGAGGACGACAAAGGTGTTCCGGGGGGAGCCACACCTACCTGTGGCAGGTG 503

Db 484 ctgaaagagaatggtccaatggcctctgaccacactgtgccttaacctactcatatcttt 543
|| || ||||| || || ||||| ||||| ||||| ||||| ||||| || || ||

Qy 504 CTTAAGGAGAACGGCCCTATGGCCAGCGACCCCTGTGCTGACCTACAGCTACCTGAGC 563

Db 544 catgtggacctggtaaaagacttgaattcaggccctcattggagccctactagtatgtaga 603
|| ||||| ||||| || || ||||| || || || || || || || || || || || || ||

Qy 564 CACGTGGACCTGTTGAAGGATCTGAACAGCGGGCTGATCGGCGCCTGTGTTGTGTCGC 623

Db 604 gaagggaagtctggccaaggaagacacagaccttgacaaaaattataactattttgtct 663
|| || || ||||| ||||| || || ||||| ||||| ||||| || || || || || || ||

Qy 624 GAGGGCAGCTGGCCAAGGAGAAAACCCAGACCCCTGCACAAGTTTCATCTCTGTGTCGCC 683

Db 664 gtatttgatgaagggaagattggcaactcagaaaaaagaactccttgatgcaggatagg 723
|| || || ||||| || ||||| || || ||||| || ||||| || ||||| || |||||

Qy 684 GTGTTGAGAGGGGAAGAGCTGGCACAGCGGCTGATCGGCGCCTGTGTTGTGTCGC 743

Db 724 gatgctgeatctgtctggtggtgccaaggaatcagtcattggcctcattgagtcagacagg 783
|| || || || ||||| || ||||| || ||||| || ||||| || || || || || || ||

Qy 744 GAGCGCGCAGCGCCCGCCTGGCCCAAGATGCACACCGTTAACCGGTACGTGAACCCG 803

Db 784 tctctgcaggtctgattggtgccaaggaatcagtcattggcctcattgagtcagtcagtc 843
||||| || ||||| || ||||| || || || ||||| || || || ||||| || || || ||

Qy 804 AGCCTGCGCGCCTGATCGGCTGCCACCGAAGAGCGGTACTTGGCACGTCATCGGCATG 863

Db 844 ggcaccaactcctgaagtgcactcaatattcctcgaaagtcacacatttcttgtgaggaac 903
||||||| ||||| ||||| || ||||| || || ||||| || || ||||| || || || || ||

Qy 864 GGCACCAACCCCTGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTGTTGGTGCAC 923

Db 904 catcgccaggcgtccttggaaatctcgccaataaacttctactgtctcaaacactcttg 963
|| ||||| || ||||| || || || || ||||| || || ||||| || || || || || ||

Qy 924 CACCGCCAGGCCAGCTGGAGATCAGCCCCCATCACCTTCCTGACTGCCAGACCTGCTG 983

Db 964 atggaccttggacagtttctactgttttgtcataatcttcccaccaaactgatggcatg 1023
||||||| || ||||| || ||||| || || || || || ||||| || || ||||| || || || ||

Qy 984 ATGGACCTAGGCCAGCTTCTGCTGTTCTGCCATATCAGAGCCACGAGCAGCGGCATG 1043

Db 1024 gaagcttatgtcaagtagacagctgtccagagggaacccccaaactacgaatgaaaaataat 1083
|| ||||| || || || ||||| || || ||||| || || || || || ||||| || || || ||

Qy 1044 GAGGCTTACGTGAAGGTGGACAGTGCCTCGAGGAGCCCGAGCTGCGCATGAAGAACAAC 1103

Db 1084 gaagaagcggaagactatgatgatcttactgatttgaatggatgggtcaggttt 1143
|| || || || ||||| || || || || || || || || || ||||| || || ||||| || ||

Qy 1104 GAGGAGCGCGAGGACTACGACGAGGACCTACCGGACAGCGAGATGGATGTCTGACGCTTC 1163

Db 1144 gatgatgacaactctccttctcctttatccaaattcgtcagttgccaaagaagatcctaaa 1203
|| || ||||| || || ||||| || || ||||| || || || ||||| ||||| |||||

Qy 1164 GACGACGACAACAGCCCCAGCTTCATCCAGATCCGACGCTGGCCAAAGAAGCACCCCTAAG 1223

Db 1204 acttgggtacattacattgctgtggaagaggagactgggactatgctcccttagtctctc 1263
|| ||||| || ||||| || || || ||||| ||||| ||||| || || ||||| || || ||

Qy 1224 ACCTGGGTGACTACATCGCGCGGAGGAGGACTGGGACTACGCCCGCTAGTACTG 1283

Db 1264 gcccccgatgacagaagttataaaagtcaatatattgaaacaatggccctcagcggattggt 1323
||||||| || || || || || || || || || || || ||||| ||||| ||||| || ||

Qy 1284 GCCCCGACGACCGCAGCTACAAGAGCCACTACCTGAACAACGGCCCCCAGCGCATCGGC 1343

Db 1324 aggaagtacaaaaagtcogatttatggcatcacagatgaaacctttaagactcgtgaa 1383
|| ||||| || || || || || || ||||| ||||| || || || || || ||||| || ||

Qy 1344 CGCAAGTACAAGAGGTGCGCTTCATGGCCTACCGCAGCAGACTTTCGAAGACCCGCGAG 1403

Db 1384 gctattcagcatgaatcaggaatcttgggacctttactttatggggaagttggagacaca 1443
|| || ||||| || || || || || || || || || || || || || || || || || || ||

Qy 1404 GCCATCCAGCACGAGTCGCGCATCTCTGGCCCCCTGCTGTACGGCGAGGTGGCGGACACC 1463

Db 1444 ctggttgattatatttaagaatacaagcaagcagaccatatataacatctaccctcacggaatc 1503
|||| |||| || || ||||| || || ||||| || || ||||| ||||| ||||| ||||| || ||

Qy 1464 CTGCTGATCATCTTCAAGAAACCGCCAGCGGCCCTACAACTCTACCCCCACGGCATC 1523

Db 1504 actgatgtccgtcctttgtattcaaggagattaccaaaaaagtgtaaaacatttgaaggat 1563
|| || || || ||||| || || ||||| ||||| || || || || || || || || |||||

Qy 1524 ACCGACGTGCGCCCCCTGTACAGCGCGCGCTGCCAAAGGCGTGAAGCACCTGAAGGAC 1583

Db 1564 tttccaattctgccaggagaaatatttcaaatataaaatggacagtgactgtagaagatggg 1623
|| || || ||||| || || || ||||| || || ||||| || || ||||| ||||| || || || ||

Qy 1584 TTCCCCATCTCTGCCCGCGGAGATCTTCAAGTACAAGTGGACCGTGCCTGAGGAGCGGC 1643

Db 1624 ccaactaaatcagatcctcggtgcctgaccgctattactctagtttcgttaatatggag 1683
|| || || || || || ||||| ||||| ||||| ||||| || || ||||| || ||||| || |||||

Qy 1644 CCCACCAAGAGCGACCCCGCTGCTTGACCGCTACTACAGCAGCTTCGTGAACATGGAG 1703

Db 1684 agagatcagcttccaggactcattggccctctcctcatctgtctacaaagaatctgtattgat 1743
|| || || || ||||| || ||||| || ||||| || || ||||| ||||| || || || || ||

Qy 1704 CCGACCTGGCCTCCGGACTGATCGGCCCCCTGCTGCATCTGCTACAAGAGACGCTGGAC 1763

Db 1744 caaagaggaaccagataaatgtcagacaaagaggaatgtcatcctgttttctgtatttgat 1803
|| || || ||||| ||||| ||||| || || || ||||| || || ||||| || || || || ||

Qy 1764 CAGCGCGCAACCCAGATCATGAGCGACAAGCGCAACGCTGATCTGTTCAAGCAGCTTCGAC 1823

Db 1804 gagaaccgaagctggtaacctcacagagaatatataaacgcttttctcccaatccagctgga 1863
||||||| ||||| ||||| || || ||||| || || ||||| || || ||||| || ||||| || |||||

Qy 1824 GAGAACCGCAGCTGGTATCTGACCGAGAACATCCAGCGCTTCTGCCCCAACCCCGCTGGC 1883

Db 1864 gtgcagcttgaggatccagagttccaagcctccaacatcatgcacagcatcaatggctat 1923
||||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1884 GTGAGCTGGAAGATCCCGAGTTCAGGGCAGCAACATCATGACAGCATCAACGGGTAC 1943

Db 1924 gtttttgatagtttgcagttgtcagtttgttggcatgaggtggcatactggtacattcta 1983
|| || || || ||||| || || || ||||| ||||| ||||| ||||| ||||| ||||| || ||

Qy 1944 GTGTTGACAGCCTGCAGCTGAGCGTGTGCTGCAFGAGGTGGCCTACTGGTACATCCTG 2003

Db 1984 agcatttggagcacagactgacttcttctgtcttcttctctgtgataaccttcaaacac 2043
||||| || || ||||| ||||| || ||||| || || ||||| || || ||||| ||||| |||||

Qy 2004 AGCATCGGCGCCAGACCCACTTCTGAGCGTGTCTTCTCGGGTATACCTTCAAGCAC 2063

Db 2044 aaaatggtctatgaagacacactcacctatttcccattctcaggagaaactgtcttcatg 2103
|| ||||| || || ||||| || ||||| ||||| ||||| || || ||||| ||||| |||||

Db	5994	tccagtggtttttagacagctggaatatgttaccatccaagaattggaattggcgggtgga	6053
Qy	3380	ccccggcggttcgacctgtgcagatgtgcgccacgaaggcgggatctggcgctgga	3439
Db	6054	atgccttatggcagcatctacatgctgggatgagcacacttttcttgtatacagcaa	6113
Qy	3440	gtgcctgatcgccgagcactgcacgcccgcgatgagcacccctgttcttggttacagcaa	3499
Db	6114	taagtgtcagactcccctgggaatggcttctggacacattagagattttcaagattacagc	6173
Qy	3500	caagtgccagacccccctggcgatggccagcgccacatccggcacttcacgatcacccg	3559
Db	6174	ttcaggacaatatggacagtggcccccagaagtggccagacttcattatcccgatcaat	6233
Qy	3560	cagcgccactacggccagtggsctcccaagctggcccgctgcactacagcggcagcat	3619
Db	6234	caatgcctggagcaccaggagcccttttcttgatacaagtgatctgttggcaccaat	6293
Qy	3620	caacgctgttcgaccaagagcccttctcttgatcaagtgagcctgctggcccccat	3679
Db	6294	gattattcaogcatcaagaccagggtgcccgtcagaagtttccagcctctacatctc	6353
Qy	3680	gatcatccaggcatacaagcccagggcgccgcagaaagttcagcagcctgtacatcag	3739
Db	6354	tcaqtttatcatcatgtatagtcttgatgggaagaadtggcagacttatcagaggaaattc	6413
Qy	3740	ccagtcatcatcatgtfactctctagacggcaagaagtggcagacactaccgcgcaacag	3799
Db	6414	cactggaaaccttaatggtcttcttggcaatgtggattcatctgggataaaaacacaatat	6473
Qy	3800	cacggcaccctgattggtgttcttggcaacgtggacagcagcgccatcaagcacacaacat	3859
Db	6474	ttttaacctccaattattgctcgatcacatccgtttgcacccaactcattatagcattcg	6533
Qy	3860	cttcaacccccccatcatcgccccgtacatccgcccgcacccccacccacacacacacac	3919
Db	6534	cagcaactcttcgcatggagttgatgggctgtgatttaaatagttgcageatgccattggg	6593
Qy	3920	cagcaccttcgcatggagctgatgggctgcgacctgaacagctgcagcatgcccttggg	3979
Db	6594	aatggagagtaagacaatatcagatgcacagattactgtcttcactcctactttacceatat	6653
Qy	3980	catgagagcaaggccatcagcgacgccccagatcacccgctccagctacttcaccaacat	4039
Db	6654	gtttgccacctggtctccttcaaaagctcgacttcacctccaagggaggagtaatgcctg	6713
Qy	4040	gttcgccaccttgagccccagcaaggcccgctgcacctgcagggccgacgaacgcctg	4099
Db	6714	gagacctcagtgtaataatccaaagagtggctgcgaagtggacttcagagaacacaatgaa	6773
Qy	4100	gccccccagctgaacaaacccccaggagtgctgcagctggacttcagaaaaaacctgaa	4159
Db	6774	agtcacaggagtaactactcagggagtaaaaatctctgcttaccagcatgtatgtgaagga	6833
Qy	4160	ggtagctggcctgacccaccagggggctcaagagcctgctgacacagcatgtacgtgaagga	4219
Db	6834	gttccctcatctccagcagtcgaagtggccatcagtggaactctctttttcagaatggcaa	6893
Qy	4220	gttccctgatcagcagcagcaggacggccacagctggaccctctgttcttccaaaaacggcaa	4279
Db	6894	agtaaaaggttttttcagggaataatcaagactccttcacacctgtgtggaactctctagacc	6953
Qy	4280	ggtagaggtgttccaggggcaaccagacacagcttcacacggctctgaacagccttgaccc	4339

```

Db      6954 accgttactgactcgtaccttcgaattaccacccagagtgggtgcaccagattgcctt 7013
          || | ||||| ||||||| || || ||||||||| ||||||| ||||| |||||
Qy     4340 CCCCCTGCTACCGGTACTTGGCATCCACCCTCAGAGCTGGTGTCACCATCGCCCT 4399

Db      7014 gaggatggagttcttggtcgaggcacaggacctctaactga 7056
          | | ||||||| ||||||| ||||| ||||||| ||||| |||||
Qy     4400 GCGCATGCAGGTCTGGCTGCGAGGCCAGGACCTGTACTGA 4442


RESULT   14
ID       T31031 standard; cDNA; 8967 BP.
AC       T31031;
DE       19-SEP-1996 (first entry)
DT       Factor-VIII full-length cDNA.
KW       Factor-VIII; retrovirus; vector; haemophilia A; gene therapy; ds.
OS       Homo sapiens.
FH       Key
FT       cds      Location/Qualifiers
              110..7165
              /*tag= a
FT       signal_peptide 110..166
              /*tag= b
FT       mat_peptide 167..7162
              /*tag= c
PN       W09621035-A2.
PD       11-JUL-1996.
PF       18-DEC-1995; U16582.
PR       30-DEC-1994; US-366851.
PA       (CHIR ) CHIRON VIAGENE INC.
PI       Bodner M, Chang S, Chi-Tang Hsu D, De Polo NJ;
DR       WPI; 96-334010/33.
DR       P-PSDB; W00465.
PT       Retroviral vector directing expression of full length factor VIII -
PT       used in the gene therapy and treatment of haemophilia A
PS       Claim 3; Page 58-68; 86pp; English.
CC       A full-length cDNA clone (T31031) codes for human Factor-VIII
CC       (W00465), a trace plasma glycoprotein which acts as a cofactor in
CC       conjunction with Factor-IXa in the activation of Factor-X.
CC       Retroviral vectors comprising the full-length cDNA can be efficiently
CC       packaged into infectious retroviral particles. These may be used to
CC       transduce cells either in vivo or ex vivo. Factor-VIII expressed
CC       from such transduced cells will be processed and transported in a
CC       fashion analogous to the expression product of a normal Factor-VIII
CC       gene. Retroviral particles harbouring such vectors will be useful
CC       in the gene therapy of haemophilia A.
SQ       Sequence 8967 BP; 2841 A; 1898 C; 1833 G; 2395 T;

Query Match      24.3%; Score 1083; DB 21; Length 8967;
Best Local Similarity 73.8%; Pred. No. 0.00e+00;
Matches 1680; Conservative 0; Mismatches 597; Indels 0; Gaps 0;

Db      167 gccaccagaagatactacctgggtgcagtggaactgtcatgggactatatgcaaatgat 226
          ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy     84 GCCACCCCGGTACTACTTGGGCGCGTGGAGCTCTCCTGGGACTACATGCACAGCGAC 143

Db      227 ctcggtgagctgctgtggacgcgaagatttcctcctagagtgccaaaatcttttcattc 286
          || | ||||| || ||||||| || | || | || ||||| || || |||||
Qy     144 CTGGGGAGCTCCCGGTGGACGCCCGCTTCCCCCGCGGTGCCAAGAGCTTCCCCCTTC 203

Db      287 aacacctcagctcgtgtacaaaaaagactctgttgtagaattcacggatcaccttttcaac 346
          ||||| | | ||||| || | ||||| || || ||||| || || ||||| |||||
Qy     204 AACACCAGCGTGGTGTACAAGAAAACCCCTGTTCGTGGAGTTACCGGACCACTGTTCAC 263

Db      347 atcgtaagccaagccaccctggatgggtctgtctaggtcctaccatccaggtcagggtt 406

```


Qy 264 ATTGCCAAGCGCGCCCCCTGGATGGCGCTGCTGGCGCCCAACCATCCAGGCGGAGGTG 323

Db 407 tatgatcacagtgtcattacacttaagaacactgggtctcccatcctgtcagtccttcagtct 466

Qy 324 TAGCACACCGTGGTGATCACCTGAAGAACATGGCCAGCCACCCCGTACGCTGCACGCC 383

Db 467 gttgggtgtatccctactggaagcttctgagggagctgaaatatgatgatcacagaccagtcaa 526

Qy 384 GTGGCGGTGAGCTACTGGAAGGCCACCGAGGGCGCCGAGTACGACGACCAAGACGTCCCGAG 443

Db 527 agggagaagaagatgataaagtcttccctggtggaagccataacatatgtctggcaggtc 586

Qy 444 CGCGAGAAGGAGGACGACAAAGGTGTTCCCGGGGGAGCCACACTACGTTGTGGCAGGTG 503

Db 587 ctgaaagagaagtgtccaatggcctctgaccacactgtgccttaacctactcataatctttct 646

Qy 504 CTTAAGGAGAACGGCCCTATGGCCACCGACCCCTGTGCCCTGACCTACAGCTACCTGAGC 563

Db 647 catgtggacctggtaaaagacttgaattcaggcctcattggagccctactagtatgtaga 706

Qy 564 CACGTGGACCTGGTGAAGGATCTGAACAGCGGGCTGATCGGGGCCCTGCTGGTGTGTGCG 623

Db 707 gaagggagctctggccaaggaagaagacacagacacttgcacaaaatttatactactttttgct 766

Qy 624 GAGGGCAGCCTGGCCHAGGAGAAAACCCAGACCCCTGCACAAATTTCATCCTGCTGTTTCGCC 683

Db 767 gtattttgatgaagggaagttggcactcagaaaacaaagaactccttgatgcaggatagg 826

Qy 684 GTGTTGCAGAGGGCAAGAGCTGGCACAGCGGACTAAGAACAGCCTGATGCAGGACCCG 743

Db 827 gatgctgeatctgctcgggcctggcctaaaaatgcacacagtcfaatggttatgtaaacagg 886

Qy 744 GACGCGCCAGCGCCCGCGCTGGCCCAAGATGCACACCGTTAACGGCTACGTGAACCCG 803

Db 887 tctctgccaggtctgattggatgccacaggaaatcagtcctatttggcatgtgattggaatg 946

Qy 804 AGCCTGCCCGCCCTGATCGGCTGCCACCGCAAGAGCGGTGTACTGGCACGTCATCGGCATG 863

Db 947 ggcaccactcctgaagtgcactcaatatcttcgaaggtcacacatacttcttgtgaggaac 1006

Qy 864 GGCACCAACCCCTGAGGTGCACAGCATCTTCTTGGAGGGCCACACTTCTTGGTGGCAAC 923

Db 1007 catcgccagcgctccttggaaatctcgccaataactttccttactgctcaaacactcttg 1066

Qy 924 CACGCGCCAGCCAGCCTGGAGATCAGCCCCCATCAACCTTCTGCTGACTGCCACAGCCCTGCTG 983

Db 1067 atggaccttggacagtttctactgttttgtcataatctcttcccaccaacatgatggcatg 1126

Qy 984 ATGACCTTAGGCCAGTTCCTGCTGTTCTGCCACATCAGCAGCCACCCAGCAGCGGCATG 1043

Db 1127 gaagcttatgtcaaatgtagacagctgtccagagggaacccccaaactacgaatgaaaaataat 1186

Qy 1044 GAGGCTTACGTGAAGGTGCACAGCTGCCCGAGGAGCCCGAGCTGCGCATGAAGAACAAC 1103

Db 1187 gaagaagcgaagactatgatgatgatcttactgattctgaaatggatgtgggtcaggttt 1246

Qy 1104 GAGGAGCCGAGGACTTACGACGACGACCTGACCGACAGCGAGATGGATGTGCTACGCTTC 1163

Db 1247 gatgatgacaactctccttcttataccaaattcgctcagttgccaaagaagcatcctaaa 1306

Qy 1164 GACGACGACAACAGGCCCCAGGTTCATCCAGATCCCGCAGCGGTGGCCAAAGAACACCCCTAAG 1223

Db 1307 acttgggtacattacattgctgctgaagaggaggactgggactatgctcccttagtcctc 1366

Qy 1224 ACCTGGGTGCACCTACATCGCCCGCGAGGAGGAGGACTGGGACTACGCCCGCTAGTACTG 1283

Db 1367 gcccccgatgacagaagttataaaagtcaatatatttgaacaatggccctcagcggaattggt 1426

Qy 1284 GCCCCCGACGACCGCGAGCTACAAAGCCAGTACCTGAAACAAGGGCCCCCAGCGCATCGGC 1343

Db 1427 aggaagtacaaaaaagtccgattttatggcatcacagatgaaacccitttaagactcgtgaa 1486

Qy 1344 CGCAAGTACAAAGAGGTGCGCTTCATGGCCCTACACCGACGAGACTTTCAAGACCCCGCGAG 1403

Db 1487 gctattcagcatgaatcaggaatcttgggaccttttactttatggggaagttggagacaca 1546

Qy 1404 GCCATCCAGCACGAGTCCGGCATCTCTGGCCCCCTGCTGTACGGCGAGGTGGCGGACACC 1463

Db 1547 ctgttgattatatttaagaatcaagcaagcagaccataataacatctaccctcacggaatc 1606

Qy 1464 CTGCTGATCATCTTCAAGAACCCAGGCCAGCGGCCCTACAAACATCTACCCCCACGGCATC 1523

Db 1607 actgatgtccgtcctttgtattcaaggaagattaccaaaaaggtgtaaaaactttgaaggat 1666

Qy 1524 ACCGACGTGCGCCCCCTGTACAGCGCGCGCTGCCAAAGGCGGTGAAGCACCTGAAGGAC 1583

Db 1667 tttccaattctgccaggagaaaaatattcaaatataaaatggacagtgactgtagaagatggg 1726

Qy 1584 TTCCCCATCCTTGCCCGCGCGAGATCTTCAAGTACAAGTGGACCGTGAACGCTGGAAGGAG 1643

Db 1727 ccaactaaatcagatcctcgggtgcctgaccgcctattactctagttttcgttaatatggag 1786

Qy 1644 CCCACCAAGAGCGCACCCCGCTGCCCTGACCCGCTACTACAGCAGCTTCGTGAACATGGAG 1703

Db 1787 agagatctagcttcaggactcctcgggtgcctgaccgcctattactctagttttcgttaatatggag 1846

Qy 1704 CGCGACCTGGCCTCCGGACTGATCGGCCCCCTGCTGATCTGCTACAAGGAGAGCGTGGAC 1763

Db 1847 caaagaggaaccagataaatgtcagacaaaggaatgtcatcctgttttctgtatttgat 1906

Qy 1764 CAGCGCGGCACCCAGATCATGAGCGACAAAGCGCAACGTCATCTGTTCAAGCGTGTTCGAC 1823

Db 1907 ggaacccgaagctggtacctcacagagaataatacaacgctttctccccaatccagctgga 1966

Qy 1824 GAGAACCGCAGCTGGTATCTGACCGAGAACATCCAGCGCTTCTGCCCAACCCCGCTGGC 1883

Db 1967 gtgcagcttgaggatccagagttccaagctccaacatcatgcacagcatcaatggctat 2026

Qy 1884 GTGCAGCTGGAAGATCCCGAGTTCGAGGCCAGCAACATCATGCAACAGCATCAACGGCTAC 1943

Db 2027 gtttttgatagtttgcagttgtcagtttgttggcatgaggtggcatactggtacattcta 2086

Qy 1944 GTGTTGCACAGCCTGCAGCTGAGCGTGTGCCCTGCATGAGGTGGCCTACTGGTACATCCCTG 2003

Db 2087 agcattggagcacagactgacttcttctgtcttcttctgtggatataaccttcaaacac 2146

Qy 2004 AGCATCGGGCCCCAGACCCGACTTCCTGAGCGCTGTTCTTCTCCGGGTATACCTTCAAGGCAC 2063

Db 2147 aaaatggtctatgaagacacactcacccctattcccattctcaggagaaaactgtcttcag 2206

Qy 2064 AAGATGGTGTACGAGGACACCCCTGACCCCTGTTCCTTCTCCGGCGGAGACTGTGTTCAATG 2123

Db 2207 tcgatggaaaaccccaggtctatggattcttggggtgccacaactcagactttcgggaacaga 2266

Qy 2124 TCTATGGAGAACCCCGCCTGTGGAATTCTTGGGCTGCCACAACAGCGGACTTCCGCCAACCCG 2183

Db 2267 ggcatagccgccttactgaaggttttctagttgtgacaagaacacactggtgattattacgag 2326

QY 2184 GGCATGACTGCCCTGCTGTAAGTCTCCAGCTGCGACAAGAACACCGGCCACTACTACGAG 2243
Db 2327 gacagttatgaagatatttcagcatacttgctgagtgaataaaacaatgccattgaaccaaga 2386
QY 2244 GACAGCTACGAGGACATCTCCGGCTFACCTGCTGTCCAAGAACAACGCCATCGAGCCCCGC 2303
Db 2387 agcttctcccagaattcaagacacccgtagcactaggcaaaaagcaatttaatgccacc 2443
QY 2304 TCCTTCTCCCAAAACTCCGCCACCCAGCAGCGCTCAGAAGCAGTTCAACGCCACC 2360

RESULT 15
ID T61548 standard; cDNA; 9009 BP.
AC T61548;
DT 03-DEC-1997 (first entry)
DE Factor VIII:C (Arg1689Lys) coding sequence.
KW Factor VIII:C; analogue; substitution; Factor VIII:C deficiency;
KW mammal; haemophilia A; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 151..7206
FT *tag= a
FT signal_peptide 151..208
FT *tag= b
FT mat_peptide 209..7203
FT *tag= c
FT /product= Factor VIII:C
FT allele 5272..5274
FT *tag= d
FT /note= "Encodes Lys"
PN W09703194-A1.
PD 30-JAN-1997.
PF 09-JUL-1996; U11441.
PR 11-JUL-1995; US-001029.
PA (CHIR) CHIRON CORP.
PI Burke RL, Rosenberg S;
DR WPI; 97-119049/11.
DR P-PSDB; W13496.
PT Factor VIII:C analogue with Arg 1689 substituted by Lys - useful for
PT preventing and treating a Factor VIII:C deficiency, particularly
PT haemophilia A
PS Claim 7; Page -; 46pp; English.
CC This sequence encodes an active Factor VIII:C polypeptide analogue.
CC The analogue comprises a native Factor VIII:C polypeptide that is
CC modified by substitution of the Arg residue at position 1689 with
CC Lys. The polypeptide analogue can be used to prevent and treat a
CC Factor VIII:C deficiency in a mammal, esp. haemophilia A. This
CC sequence is not given in the specification and is based on the
CC sequence given in Genbank Ref. K01740.
SQ Sequence 9009 BP; 2855 A; 1905 C; 1843 G; 2405 T;

Query Match 24.3%; Score 1083; DB 33; Length 9009;
Best Local Similarity 73.8%; Pred. No. 0.00e+00;
Matches 1680; Conservative 0; Mismatches 597; Indels 0; Gaps 0;

Db 208 gccaccagaagatactacctgggtgcagtggaactgtcatgggactatatgcaaaagtgat 267
QY 84 GCCACCGCCCGCTACTACCTGGGGCGCGTGGAGCTGTCTGGGACTACATGCGAGCGGAC 143
Db 268 ctcggtgagctgcctgtggacgcaagatttccctcctagagtgcaaaaattctttccattc 327
QY 144 CTGGCGGAGCTCCCGGTGGAGCCCGCTTCCCCCCCCCGGTGCCCAAGAGCTTCCCC TTC 203

Db 328 aacacctcagtcgtgtacaaaaagactctgtttgtagaattcacggttcaccttttcaac 387
QY 204 AACACCAGCGTGGTGTACAAGAAAACCTGTTCGTGGAGTTCACCGACCACCTGTTCAAC 263
Db 388 atcgctaagccaagccaccctggatgggtctctgtaggtcctaccatccaggctgaggtt 447
QY 264 ATTGCCAAGCCGGCCCCCTCGGATGGGCCTGCTGGGCCCCACCATCCAGGCCGAGGTG 323
Db 448 tatgatacagtggtcattacacttaagaacatggcttcccatcctgtcagtccttcatgct 507
QY 324 TACGACACCGTGGTGATCACCCCTGAAGAACATGGCCAGGCCACCCCGTCAGCCGTGCACGCC 383
Db 508 gttggtgtatcctactggaagccttctgagggaggtgaaatatgatgatcagaccagtcaa 567
QY 384 GTGGCGTGAGCTACTGGAAGGCCAGCGGAGGGCGCGGAGTACGACGACGAGCTCCAG 443
Db 568 agggagaagaagatgataaagtcttccctggtggaagccatacatatgtctggcaggtc 627
QY 444 CCGGAGAAGGAGGACGACAAAGGTGTTCCCGGGGGGAGGCCACACCTACGTTGGCAGGTG 503
Db 628 ctgaaagagaatggtccaatggcctctgacccactgtgccttacctactcatatctttct 687
QY 504 CTTAAGGAGAACGGCCCTATGGCCAGCGACCCCTGTGCTGACCTACAGCTACCTGAGC 563
Db 688 catgtggacctggtaaaagacttgaattcaggcctcattggagccctactactagtatgtaga 747
QY 564 CACGTGCACCTGGTGAAGGATCTGAACAGCGGGGTGATCGGGGCCCTGCTGGTGTGTCGC 623
Db 748 gaagggagtcctggccaaggaagacacagaccttgcaaaatttatactacttttgc 807
QY 624 GAGGGCAGCCTGGCCACGAGAGAAAACCCAGACCCCTGCACAAGTTTCCTCTGCTGTCGCC 683
Db 808 gtatttgatgaagggaagtgtggcactcagaaacaaagaactccttgatgcaggatagg 867
QY 684 GTGTTCCAGCAGGGGAAGAGCTGGCACAGCGAGACTAAGAACAGCCCTGATGCAGGACCGC 743
Db 868 gatgtgcatctgctcgggctggcctaaagtgcacacagtcagtgttatgtaaacagg 927
QY 744 GACGCCGCCAGCCCGCGCCTGGCCCCAAGATGCACACCCGTTAACGGCTACGTGAACCCG 803
Db 928 tctctgccaggtctgattggatgccacaggaatacagtcctattggcatgtgattggaatg 987
QY 804 AGCCTGCCCGCGCTGATCGGCTGCCACCCCAAGACCGCTGTACTGGCACCGTCATCGGCATG 863
Db 988 ggcaccactctgaagtgcactcaatatctctcggaaggtcacacatttctgtgaggaac 1047
QY 864 GGCACACCCCTGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTCTGGTGGCGAAC 923
Db 1048 catcgccaggcgtccttggaatatctcgccaataaacttcttactgctcaaacactcttg 1107
QY 924 CACCGCCAGGCCAGCCTGGAGATCAGCCCCCATCACCTTCCTGACTGCCCCAGACCCCTGCTG 983
Db 1108 atggaccttggacagtttctactgttttgtcatactcttcccaccaacatgatggcatg 1167
QY 984 ATGGACCTAGGCCAGTTCTCTGCTGTTCTGCCACATCAGCAGCCACCCAGCAGCGGCATG 1043
Db 1168 gaagcttatgtcaagtagacagctgtccagagggaaccccccaactacgaatgaaaaaat 1227
QY 1044 GAGGCTTACGTGAAGGTGGACAGCTGCCCGGAGGAGCCCCCAGCTGCCATGAAGACAAC 1103
Db 1228 gaagaagcgggaagactatgatgatgacttactgattctgaaatggatgtggtcaggttt 1287
QY 1104 GAGGAGCCCGAGACTACGACGAGGACCTTGACCGGAGCGGAGATGGATGTCGTACGCTTC 1163

